

GenCore version 5.1.3
Copyright (c) 1993 - 2002 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: October 29, 2002, 09:25:41 ; Search time 20.636 Seconds
(without alignments)
2556.865 Million cell updates/sec

Title: US-08-325-278B-1
Perfect score: 1565
Sequence: 1 AVENKEETPTETDSEEV.....GGYTINRFAGKKVDERPEE 305

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 562222 seqs, 172994929 residues

Total number of hits satisfying chosen parameters: 562222

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : SPTREMBL19.*

- 1: sp_archaea.*
- 2: sp_bacteria.*
- 3: sp_fungi.*
- 4: sp_human.*
- 5: sp_invertebrate.*
- 6: sp_mammal.*
- 7: sp_mhc.*
- 8: sp_organelle.*
- 9: sp_phase.*
- 10: sp_plant.*
- 11: sp_rodent.*
- 12: sp_virus.*
- 13: sp_vertebrate.*
- 14: sp_unclassified.*
- 15: sp_rvirus.*
- 16: sp_bacteriaph.*
- 17: sp_archaeap.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Match	Query Length	ID	Description
1	1565	100.0	455	2 Q53291	Q53291 streptococc
2	1561	99.7	719	2 Q51912	Q51912 peptostrept
3	1226.5	78.4	992	2 Q51918	Q51918 peptostrept
4	167	10.7	664	2 Q53975	Q53975 streptococc
5	166	10.6	479	10 Q9LF88	Q9LF88 arabidopsis
6	166	10.6	669	2 Q93EM8	Q93EM8 streptococc
7	151	9.6	448	10 Q9SKP0	Q9SKP0 arabidopsis
8	149	9.5	448	10 Q96246	Q96246 arabidopsis
9	148.5	9.5	2276	2 Q93TY6	Q93TY6 streptococc
10	141.5	9.0	1110	13 Q91255	Q91255 streptococc
11	141	9.0	1849	2 Q93AK2	Q93AK2 streptococc
12	137.5	8.8	992	16 Q9WZAG	Q9WZAG thermotoga
13	137.5	8.8	5458	5 Q9U459	Q9U459 plasmodium
14	136	8.7	506	2 Q53837	Q53837 salmonella
15	134.5	8.6	3198	5 Q9U8G8	Q9U8G8 manduca sex
16	132.5	8.5	737	16 Q9JYK4	Q9JYK4 neisseria m

17	132.5	8.5	17352	5 Q95YM2	Q95YM2 procambarus
18	131.5	8.4	891	2 Q04111	Q04111 enterococcu
19	131.5	8.4	1166	2 Q86489	Q86489 staphylococ
20	130.5	8.3	564	2 Q89188	Q89188 streptococ
21	130.5	8.3	581	2 Q33742	Q33742 streptococ
22	130.5	8.3	663	2 Q30874	Q30874 streptococ
23	130.5	8.3	701	2 Q9KK48	Q9KK48 streptococ
24	130.5	8.3	1141	16 Q9W46	Q9W46 staphylococ
25	130.5	8.3	1141	16 Q932F7	Q932F7 staphylococ
26	129.5	8.3	1733	2 Q8K114	Q8K114 staphylococ
27	129	8.2	1171	2 Q9KW6	Q9KW6 staphylococ
28	128.5	8.2	938	16 Q927R4	Q927R4 listeria in
29	128.5	8.2	1119	13 P87344	P87344 thesagra ch
30	127.5	8.1	701	2 Q9R0T5	Q9R0T5 streptococ
31	127.5	8.1	717	2 Q9EY85	Q9EY85 streptococ
32	127.5	8.1	728	2 Q924N2	Q924N2 clostridium
33	127.5	8.1	843	2 Q47802	Q47802 neisseria g
34	127	8.1	913	13 Q13099	Q13099 xenopus lae
35	126.5	8.1	1035	5 Q21380	Q21380 caenorhabdi
36	126.5	8.1	2045	16 Q9AOK5	Q9AOK5 streptococ
37	126	8.1	955	2 Q45574	Q45574 bacillus sp
38	125.5	8.0	673	10 Q9M955	Q9M955 arabidopsis
39	125.5	8.0	1510	5 Q25920	Q25920 plasmodium
40	125	8.0	463	10 Q39871	Q39871 glycine max
41	125	8.0	2083	5 Q9N435	Q9N435 caenorhabdi
42	124.5	8.0	1302	2 Q49547	Q49547 mycoplasma
43	124	7.9	495	16 Q9M9G8	Q9M9G8 staphylococ
44	124	7.9	506	2 Q53834	Q53834 salmonella
45	124	7.9	1489	10 Q96449	Q96449 phytophthor

ALIGNMENTS

RESULT 1

Q53291 ID Q53291 PRELIMINARY; PRT; 455 AA.
AC Q53291;
DT 01-NOV-1996 (TREMBLrel. 01, Created)
DT 01-NOV-1996 (TREMBLrel. 01, Last sequence update)
DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)
DE PROTEIN LG (FRAGMENT).
OS Streptococcus sp.
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Streptococcaceae;
OC Streptococcus.
OX NCBI_TaxID=1306;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=93094283; PubMed=1460053;
RA Kihlberg B.M., Sjöbring U., Kastern W., Björck L.;
RT "Protein LG: a hybrid molecule with unique immunoglobulin binding
RT properties";
RL J. Biol. Chem. 267:25583-25588(1992). Dec 1992
DR EMBL; SS0809; AAA03280.1; -
DR HSSP; P06654; LPGA.
DR InterPro; IPR003147; B1.
DR InterPro; IPR000724; IGG_bind_B.
DR Pfam; PF02246; B1; 4.
DR Pfam; PF01378; IGG_binding_B; 2.
FT NON_TER 455 455
SQ SEQUENCE 455 AA; 49926 MW; 381FC235BBC8307B CRC64;

Query Match 100.0%; Score 1565; DB 2; Length 455;
Best Local Similarity 100.0%; Pred. No. 9, le-79;
Matches 305; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 AVENKEETPTETDSEEV...TANLIFANGSTQTEFEKGTPEKATSEAYAYADTLKDN 60
Db 22 AVENKEETPTETDSEEV...TANLIFANGSTQTEFEKGTPEKATSEAYAYADTLKDN 81
QY 61 GEYTVDAVADKGYTLNFKAGKTEPEEPKEEVTIKANLIYADGKTOTAEFKGFEEATAE 120
Db 82 GEYTVDAVADKGYTLNFKAGKTEPEEPKEEVTIKANLIYADGKTOTAEFKGFEEATAE 141

THIS PAGE BLANK (uspto)

QY 121 AYRYADALKKNGEYTVDVADKGYTLNIFAGKEKTPPEPKEEVTIKANLIYADGKTQTA 180
 Db 142 AYRYADALKKNGEYTVDVADKGYTLNIFAGKEKTPPEPKEEVTIKANLIYADGKTQTA 201
 QY 181 EKGTFEEATAEAYRYADLLAKENGKYTVDVADKGYTLNIFAGKEKTPPEPKEEVTIKA 240
 Db 202 EKGTFEEATAEAYRYADLLAKENGKYTVDVADKGYTLNIFAGKEKTPPEPKEEVTIKA 261
 QY 241 NLIYADGKTQTAEFKGTFAEATAEAYRYADLLAKENGKYTVADLEGGYTNIRFAGKKYD 300
 Db 262 NLIYADGKTQTAEFKGTFAEATAEAYRYADLLAKENGKYTVADLEGGYTNIRFAGKKYD 321
 QY 301 EKPEE 305
 Db 322 EKPEE 326

RESULT 2

Q51912 PRELIMINARY; PRT; 719 AA.
 ID Q51912
 AC Q51912;
 DT 01-NOV-1996 (TReMBLrel. 01, Created)
 DT 01-NOV-1996 (TReMBLrel. 01, Last sequence update)
 DT 01-DEC-2001 (TReMBLrel. 19, Last annotation update)
 DE PROTEIN L PRECURSOR.
 OS Peptostreptococcus magnus.
 OC Bacteria; Firmicutes; Bacillus/Clostridium group; Clostridiaceae;
 OC Finegoldia.
 OX NCBI_TaxID=1260;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=312;
 RX MEDLINE=90215984; PubMed=2108927;
 RA Kasteren W., Holst E., Nielsen E., Sjoerding U., Bjorck L.;
 RT "Protein L, a bacterial immunoglobulin-binding protein and possible
 RT virulence determinant.";
 RL Infect. Immun. 58:1217-1222(1990).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=312;
 RX MEDLINE=92316971; PubMed=1618782;
 RA Bjoerck L., Sjoerding U., Kasteren W.;
 RT "Structure of peptostreptococcal protein L and identification of
 RT repeated immunoglobulin light chain-binding domain.";
 RL J. Biol. Chem. 267:12820-12825(1992).
 DR EMBL; M86697; AAA25612.1;
 DR InterPro; IPR003147; B1.
 DR InterPro; IPR001899; Gram_pos_anchor.
 DR Pfam; PF02246; B1; 5.
 DR Pfam; PF00746; Gram_pos_anchor; 1.
 KW Signal.
 FT SIGNAL
 FT CHAIN
 SQ SEQUENCE 719 AA; 78983 MW; 963A8D76D5E34DD2 CRC64;

Query Match 99.7%; Score 1561; DB 2; Length 719;
 Best Local Similarity 100.0%; Pred. No. 2.5e-78;
 Matches 304; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 VENKEETPETDSEEEVTIKANLIYADGKTQTAEFKGTFAEATAEAYRYADLLAKKNG 61
 Db 95 VENKEETPETDSEEEVTIKANLIYADGKTQTAEFKGTFAEATAEAYRYADLLAKKNG 154
 QY 62 EYTVADKGYTLNIFAGKEKTPPEPKEEVTIKANLIYADGKTQTAEFKGTFAEATAE 121
 Db 155 EYTVADKGYTLNIFAGKEKTPPEPKEEVTIKANLIYADGKTQTAEFKGTFAEATAE 214
 QY 122 YRYADALKKNGEYTVDVADKGYTLNIFAGKEKTPPEPKEEVTIKANLIYADGKTQTA 181
 Db 155 YRYADALKKNGEYTVDVADKGYTLNIFAGKEKTPPEPKEEVTIKANLIYADGKTQTA 274
 QY 182 FKGTFEEATAEAYRYADLLAKENGKYTVDVADKGYTLNIFAGKEKTPPEPKEEVTIKAN 241

Db 275 FKGTFEEATAEAYRYADLLAKENGKYTVDVADKGYTLNIFAGKEKTPPEPKEEVTIKAN 334
 QY 242 LIYADGKTQTAEFKGTFAEATAEAYRYADLLAKENGKYTVADLEGGYTNIRFAGKKYD 301
 Db 335 LIYADGKTQTAEFKGTFAEATAEAYRYADLLAKENGKYTVADLEGGYTNIRFAGKKYD 394
 QY 302 KPEE 305
 Db 395 KPEE 398

RESULT 3

Q51918 PRELIMINARY; PRT; 992 AA.
 ID Q51918
 AC Q51918;
 DT 01-NOV-1996 (TReMBLrel. 01, Created)
 DT 01-NOV-1996 (TReMBLrel. 01, Last sequence update)
 DT 01-DEC-2001 (TReMBLrel. 19, Last annotation update)
 DE PROTEIN L PRECURSOR.
 OS Peptostreptococcus magnus.
 OC Bacteria; Firmicutes; Bacillus/Clostridium group; Clostridiaceae;
 OC Finegoldia.
 OX NCBI_TaxID=1260;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=3316;
 RX MEDLINE=95078460; PubMed=7987012;
 RA Murphy J.P., Trowern A.R., Dugleby C.J.;
 RT "Nucleotide sequence of the gene for peptostreptococcal protein L.";
 RL DNA Seq. 4:259-265(1994).
 DR EMBL; L04466; AAA67503.1;
 DR HSSP; Q51911; IGAB.
 DR InterPro; IPR003147; B1.
 DR InterPro; IPR002988; GA.
 DR InterPro; IPR001899; Gram_pos_anchor.
 DR Pfam; PF02246; B1; 4.
 DR Pfam; PF01468; GA; 4.
 DR Pfam; PF00746; Gram_pos_anchor; 1.
 KW Signal.
 FT SIGNAL
 FT CHAIN
 SQ SEQUENCE 992 AA; 108700 MW; 9CFF5771578A5DCE CRC64;

Query Match 78.4%; Score 1226.5; DB 2; Length 992;
 Best Local Similarity 79.9%; Pred. No. 8.5e-60;
 Matches 246; Conservative 22; Mismatches 31; Indels 9; Gaps 4;

QY 2 VENKEETPETDSEEEVTIKANLIYADGKTQTAEFKGTFAEATAEAYRYADLLAKKNG 61
 Db 241 MERKLESEKTEP--PEEEVTIKANLIYADGKTQTAEFKGTFAEATAEAYRYADLLAKKNG 298
 QY 62 EYTVADKGYTLNIFAGKEKTPPEPKEEVTIKANLIYADGKTQTAEFKGTFAEATAE 121
 Db 299 EYTVADKGYTLNIFAGKEKTPPEPKEEVTIKANLIYADGKTQTAEFKGTFAEATAE 358
 QY 122 YRYADALKKNGEYTVDVADKGYTLNIFAGKE--KTPPEPKEEVTIKANLIYADGKTOT 179
 Db 359 YAYADLLAKENGKYTVADLEGGYTNIRFAGKE--KTPPEPKEEVTIKANLIYADGKTOT 418
 QY 180 AEFKGTFAEATAEAYRYADLLAKENGKYTVDVADKGYTLNIFAGKE--KTPPEPKEEVT 237
 Db 419 AEFKGTFAEATAEAYRYADLLAKENGKYTVADLEGGYTNIRFAGKE--KTPPEPKEEVT 478
 QY 238 IKANLIYADGKTQTAEFKGTFAEATAEAYRYADLLAKENGKYTVADLEGGYTNIRFAGK 297
 Db 479 IKVNLIFADGKTQTAEFKGTFAEATAEAYRYADLLAKENGKYTVADLEGGYTNIRFAGK 538
 QY 298 KVDKPEE 305
 Db 539 ---EOPGE 543

THIS PAGE BLANK (USPTO)

GenCore version 5.1.3
Copyright (c) 1993 - 2002 Compugen Ltd.

OM protein - protein search, using sw model

Run on: October 29, 2002, 09:25:41 ; Search time 29.364 seconds
(without alignments)
2556.865 Million cell updates/sec

Title: US-08-325-278B-3
Perfect score: 2235
Sequence: 1 AVENKEETPETDSEEV.....GVDGWTYDDATKFTVTM 434

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 562222 seqs, 172994929 residues
Total number of hits satisfying chosen parameters: 562222

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : SPREMBL_19:*

- 1: sp_archaea:*
- 2: sp_bacteria:*
- 3: sp_fungi:*
- 4: sp_human:*
- 5: sp_invertebrate:*
- 6: sp_mammal:*
- 7: sp_mhc:*
- 8: sp_organelle:*
- 9: sp_phase:*
- 10: sp_plant:*
- 11: sp_rodent:*
- 12: sp_virus:*
- 13: sp_vertebrate:*
- 14: sp_unclassified:*
- 15: sp_rvirus:*
- 16: sp_bacteriaph:*
- 17: sp_archaeap:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	ID	Description
1	2235	100.0	455	2 Q53291	Q53291 streptococc
2	1587.5	71.0	719	2 Q51912	Q51912 peptostrept
3	1263.5	56.5	992	2 Q51918	Q51918 peptostrept
4	706.5	31.6	664	2 Q53975	Q53975 streptococc
5	697.5	31.2	669	2 Q938M8	Q938M8 streptococc
6	670	30.0	185	2 Q54181	Q54181 streptococc
7	481	21.5	429	2 Q56212	Q56212 streptococc
8	404	18.1	413	2 Q53974	Q53974 streptococc
9	294	13.2	60	2 Q53337	Q53337 streptococc
10	182.5	8.2	3198	5 Q9UG88	Q9UG88 manduca sex
11	182	8.1	479	10 Q9LF88	Q9LF88 arabidopsis
12	177.5	7.9	2402	2 Q9AER7	Q9AER7 staphylococ
13	175	7.8	448	10 Q9SKP0	Q9SKP0 arabidopsis
14	170.5	7.6	1849	2 Q9SAK2	Q9SAK2 lactobacill
15	170	7.6	2276	2 Q93TY6	Q93TY6 staphylococ
16	168	7.5	448	10 Q96246	Q96246 arabidopsis

17	162.5	7.3	2353	2	P71401	P71401 haemophilus
18	162	7.2	913	13	O13099	O13099 xenopus lae
19	161	7.2	2692	4	Q9Y503	Q9Y503 homo sapien
20	160.5	7.2	5458	5	Q9U459	Q9U459 plasmodium
21	160	7.2	2691	4	Q9NS12	Q9NS12 homo sapien
22	160	7.2	2705	4	Q9Y503	Q9Y503 homo sapien
23	160	7.2	2725	4	Q9NYE5	Q9NYE5 homo sapien
24	159	7.1	1856	16	Q97T80	Q97T80 streptococc
25	158	7.1	2045	16	Q9AOK5	Q9AOK5 streptococc
26	157.5	7.0	1166	2	O86489	O86489 staphylococ
27	157	7.0	798	13	Q90307	Q90307 carassius a
28	156.5	7.0	584	2	O69188	O69188 streptococc
29	156.5	7.0	581	2	O33742	O33742 streptococc
30	156.5	7.0	663	2	O30874	O30874 streptococc
31	156.5	7.0	701	2	O9KK48	O9KK48 streptococc
32	156.5	7.0	969	5	Q9NDI9	Q9NDI9 plasmodium
33	156	7.0	1368	2	Q93M90	Q93M90 clostridium
34	154.5	6.9	1110	13	Q91255	Q91255 petromyzon
35	154.5	6.9	1822	2	Q07290	Q07290 streptococc
36	154	6.9	737	16	Q9JYK4	Q9JYK4 neisseria m
37	154	6.9	845	5	Q9Y1P8	Q9Y1P8 plasmodium
38	154	6.9	1698	2	O9LC00	O9LC00 staphylococ
39	153.5	6.9	492	2	Q56193	Q56193 staphylococ
40	153.5	6.9	573	2	Q56192	Q56192 staphylococ
41	153.5	6.9	701	2	Q9RQT5	Q9RQT5 streptococc
42	153.5	6.9	1873	2	Q9Z4N7	Q9Z4N7 enterococcu
43	153	6.8	897	13	O13098	O13098 xenopus lae
44	152	6.8	3147	5	Q17464	Q17464 caenorhabdi
45	151.5	6.8	717	2	Q9EY85	Q9EY85 clostridium

ALIGNMENTS

RESULT 1

ID	Q53291	PRELIMINARY;	PRT;	455 AA.
AC	Q53291;			
DT	01-NOV-1996 (T-EMBLrel. 01, Created)			
DT	01-NOV-1996 (T-EMBLrel. 01, Last sequence update)			
DT	01-DEC-2001 (T-EMBLrel. 19, Last annotation update)			
DE	PROTEIN LG (FRAGMENT).			
OS	Streptococcus sp.			
OC	Bacteria; Firmicutes; Bacillus/Clostridium group; Streptococcaceae;			
OC	Streptococcus			
OX	NCBI_TaxID=1306;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RX	MEDLINE=93094283; PubMed=1460053;			
RA	Kihlberg B.M., Sjobring U., Kastern W., Bjorck L.;			
RT	"Protein LG: a hybrid molecule with unique immunoglobulin binding			
RT	properties."			
RL	J. Biol. Chem. 267:25583-25588(1992).			
DR	EMBL; S50809; AAA03280.1; -.			
DR	HSPP; P06654; 1PGX.			
DR	InterPro; IPR003147; BL.			
DR	InterPro; IPR00724; IgG_bind_B.			
DR	Pfam; PF02246; BL; 4.			
DR	Pfam; PF01378; IgG_binding_B; 2.			
FT	NON_TER 455 455			
SQ	SEQUENCE 455 AA; 49926 MW; 381FC235BBC8307B CRC64;			

Query Match	100.0%;	Score	2235;	DB 2;	Length	455;	
Best Local Similarity	100.0%;	Pred. No.	1e-100;				
Matches	434;	Conservative	0;	Mismatches	0;	Gaps	0;
Qy	1	AVENKEETPETDSEEVTKANLIFANGSTOTAEFKGTFEKATSEAYADTLKKDN	60				
Db	22	AVENKEETPETDSEEVTKANLIFANGSTOTAEFKGTFEKATSEAYADTLKKDN	81				
Qy	61	GEYTVDVADKGYTLNFKAGKEKTEPEEKVEVTKANLIYADGKTQTATAEFKGTFEEATAE	120				
Db	82	GEYTVDVADKGYTLNFKAGKEKTEPEEKVEVTKANLIYADGKTQTATAEFKGTFEEATAE	141				

```
QY 121 AYRYADALKKDNGEYTVDVADKGYTLNKFAGKEKTPPEPKKEEVTIKANLIYADGKTQTA 180
|||||
Db 142 AYRYADALKKDNGEYTVDVADKGYTLNKFAGKEKTPPEPKKEEVTIKANLIYADGKTQTA 201
|||||
QY 181 EFKGTPEEATAEAYRYADLLAKENGYTVDVADKGYTLNKFAGKEKTPPEPKKEEVTIKA 240
|||||
Db 202 EFKGTPEEATAEAYRYADLLAKENGYTVDVADKGYTLNKFAGKEKTPPEPKKEEVTIKA 261
|||||
QY 241 NLIYADGKTQTAEFKGTFAEATAEAYRYADLLAKENGYTADLEDGGYTNIRFAGKKVD 300
|||||
Db 262 NLIYADGKTQTAEFKGTFAEATAEAYRYADLLAKENGYTADLEDGGYTNIRFAGKKVD 321
|||||
QY 301 EKPEEPMDTYKLLILNGKTLKGETTTEAVDAATAEKVFKQYANDNGVDGEWYTDATKTFT 360
|||||
Db 322 EKPEEPMDTYKLLILNGKTLKGETTTEAVDAATAEKVFKQYANDNGVDGEWYTDATKTFT 381
|||||
QY 361 VTEKPEVIDASELTPAVTTYKLVINGKTLKGETTTTAVDAETAFAKQYANDNGVDGVW 420
|||||
Db 382 VTEKPEVIDASELTPAVTTYKLVINGKTLKGETTTTAVDAETAFAKQYANDNGVDGVW 441
|||||
QY 421 TYDDATKTFTVTEM 434
|||||
Db 442 TYDDATKTFTVTEM 455
|||||
```

```
RESULT 2
Q51912
ID Q51912 PRELIMINARY; PRT; 719 AA.
AC Q51912;
DT 01-NOV-1996 (TrEMBLrel. 01, Created)
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE PROTEIN L PRECURSOR.
OS Peptostreptococcus magnus.
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Clostridiaceae;
OC Finegoldia.
OX NCBI_TaxID=1260;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=312;
RX MEDLINE=90215984; PubMed=2108927;
RA Kastern W., Holst E., Nielsen E., Sjoebing U., Bjorck L.;
RT "Protein L, a bacterial immunoglobulin-binding protein and possible
RL virulence determinant.";
RN Infect. Immun. 58:1217-1222(1990).
RP SEQUENCE FROM N.A.
RC STRAIN=312;
RX MEDLINE=92316971; PubMed=1618782;
RA Bjoerck L., Sjoebing U., Kastern W.;
RT "Structure of peptostreptococcal protein L and identification of
RL repeated immunoglobulin light chain-binding domain.";
RN J. Biol. Chem. 267:12820-12825(1992).
DR EMBL; M86697; AAA25612.1; -.
DR InterPro; IPR003147; B1.
DR InterPro; IPR001899; Gram_pos_anchor.
DR Pfam; PF02246; B1; 5.
DR Pfam; PF00746; Gram_pos_anchor; 1.
KW Signal.
FT SIGNAL.
FT CHAIN 1 18 POTENTIAL.
FT CHAIN 19 719 PROTEIN L.
SQ SEQUENCE 719 AA; 78983 MW; 963A8D76D5E34DD2 CRC64;
```

Query Match 71.0%; Score 1587.5; DB 2; Length 719;
Best Local Similarity 76.3%; Pred. No. 3e-69;
Matches 334; Conservative 13; Mismatches 56; Indels 35; Gaps 6;

```
QY 2 VENKEETPETPDSEEVTIKANLIFANGSTQTAEFKGTFAEAYAYADTLKKDNG 61
|||||
Db 95 VENKEETPETPDSEEVTIKANLIFANGSTQTAEFKGTFAEAYAYADTLKKDNG 154
|||||
QY 62 EYTVDVADKGYTLNKFAGKEKTPPEPKKEEVTIKANLIYADGKTQTAEFKGTFAEATAEA 121
```

```
Db 155 EYTVDVADKGYTLNKFAGKEKTPPEPKKEEVTIKANLIYADGKTQTAEFKGTFAEATAEA 214
|||||
QY 122 YRYADALKKDNGEYTVDVADKGYTLNKFAGKEKTPPEPKKEEVTIKANLIYADGKTQTA 181
|||||
Db 215 YRYADALKKDNGEYTVDVADKGYTLNKFAGKEKTPPEPKKEEVTIKANLIYADGKTQTA 274
|||||
QY 182 FKGTFPEATAEAYRYADLLAKENGYTVDVADKGYTLNKFAGKEKTPPEPKKEEVTIKAN 241
|||||
Db 275 FKGTFPEATAEAYRYADLLAKENGYTVDVADKGYTLNKFAGKEKTPPEPKKEEVTIKAN 334
|||||
QY 242 LIYADGKTQTAEFKGTFAEATAEAYRYADLLAKENGYTADLEDGGYTNIRFAGKKVDE 301
|||||
Db 335 LIYADGKTQTAEFKGTFAEATAEAYRYADLLAKENGYTADLEDGGYTNIRFAGKKVDE 394
|||||
QY 302 KPEEPMD-TYK--LIUNGKTLKGETTTEAVDAATAEKVFKQYANDNGVD-GEWYTDATK 357
|||||
Db 395 KPEEREQVTIKENIYEDGTQVATFKGTFAEATAEAY--RYADLLSKEHGYTADLEDG 452
|||||
QY 358 TPTVT-----EKPEVIDASELTPAVTTYKLVINGKTLKGETTTTAVDAETAFAE 404
|||||
Db 453 GYTNIRFAGKEPEETPEKPEVOD-----GYASYEEAAAKAEALKND 496
|||||
QY 405 KAFKQYANDNGVDGVWY 422
|||||
Db 497 DVNKSYYTIRGADGRYY 514
|||||
```

```
RESULT 3
Q51918
ID Q51918 PRELIMINARY; PRT; 992 AA.
AC Q51918;
DT 01-NOV-1996 (TrEMBLrel. 01, Created)
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE PROTEIN L PRECURSOR.
OS Peptostreptococcus magnus.
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Clostridiaceae;
OC Finegoldia.
OX NCBI_TaxID=1260;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=3316;
RX MEDLINE=95078460; PubMed=7987012;
RA Murphy J.P., Trovern A.R., Duggleby C.J.;
RT "Nucleotide sequence of the gene for peptostreptococcal protein L.";
RL DNA Seq. 4:259-265(1994).
DR EMBL; L04466; AAA67503.1; -.
DR HSSP; Q51911; 1GAB.
DR InterPro; IPR003147; B1.
DR InterPro; IPR002988; GA.
DR InterPro; IPR001899; Gram_pos_anchor.
DR Pfam; PF02246; B1; 4.
DR Pfam; PF01468; GA; 4.
DR Pfam; PF00746; Gram_pos_anchor; 1.
KW Signal.
FT SIGNAL.
FT CHAIN 1 24 POTENTIAL.
FT CHAIN 25 992 PROTEIN L.
SQ SEQUENCE 992 AA; 108700 MW; 9CFF571578A5DCE CRC64;
```

Query Match 56.5%; Score 1263.5; DB 2; Length 992;
Best Local Similarity 58.8%; Pred. No. 1.8e-53;
Matches 281; Conservative 36; Mismatches 90; Indels 71; Gaps 12;

```
QY 2 VENKEETPETPDSEEVTIKANLIFANGSTQTAEFKGTFAEAYAYADTLKKDNG 61
|||||
Db 241 MERKLEKETPE--PEEEVTIKANLIFANGSTQNAEFKGTFAKAVSDAYADALKKNG 298
|||||
QY 62 EYTVDVADKGYTLNKFAGKEKTPPEPKKEEVTIKANLIYADGKTQTAEFKGTFAEATAEA 121
|||||
Db 299 EYTVDVADKGYTLNKFAGKEKTPPEPKKEEVTIKVNLIFADGKTQTAEFKGTFAEATAKA 358
|||||
QY 122 YRYADALKKDNGEYTVDVADKGYTLNKFAGKE--KTPEEPKEEVTIKANLIYADGKTQ 179
|||||
```


[illegible]

```
Db 141 DAASVDQVNAAIN-----DAHTAIADITGA---ALLAKEAAAINELKQYGISDYVVTLINK 193
QY 287 GYTI-----NIRFAGKVKDEKPEP-MDTYKLIINGKTLKGETTEAADAATAE 334
Db 194 KTVGEGNALKAEILSALSPSEVIDAELTPALTSKVLKIGATFSGETATKAVDRAVAE 253
QY 335 KVFQYANDNGVDEYTDATKFTVTPEK--EVIDASELTPAVTTYKLIVNGKTLKGE 392
Db 254 QTPRDYANKNGVDGVWAYDAATKFTVTPEQVAETIEAAELTPALTYRLVIGKVTFSGE 313
QY 393 TTTKAVDAETAFAKQYANDNGVDGWYDDATKFTVTTE 433
Db 314 TATKAVDAATAEAQFRQYANDNGVTGEWAYDAATKFTVTTE 354

RESULT 8
Q53974
ID Q53974 PRELIMINARY; PRT; 413 AA.
AC Q53974;
DT 01-NOV-1996 (TREMBLrel. 01, Created)
DT 01-NOV-1996 (TREMBLrel. 01, Last sequence update)
DE MAG PROTEIN PRECURSOR.
GN MAG.
OS Streptococcus dysgalactiae.
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Streptococcaceae;
OC Streptococcus.
OX NCBI_TaxID=1334;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=94259307; PubMed=7515368;
RA Jonsson H., Frykberg L., Rantamaeki L., Guss B.;
RT "MAG, a novel plasma protein receptor from Streptococcus
dysgalactiae.";
RL Gene 143:85-89(1994).
DR EMBL; L27798; AAA26921.1; -.
DR HSSP; P06654; LPGX.
DR InterPro; IPR002988; GA.
DR InterPro; IPR001899; Gram_pos_anchor.
DR InterPro; IPR000724; IgG_bind_B.
DR Pfam; PF01468; GA; 1.
DR Pfam; PF00746; Gram_pos_anchor; 1.
DR Pfam; PF01378; IgG_binding_B; 1.
DR PRINTS; PR00015; GPOSANCHOR.
DR PROSITE; PS00343; GRAM_POS_ANCHORING; UNKNOWN_1.
KW Signal; Transmembrane.
FT SIGNAL 1 34 POTENTIAL.
FT CHAIN 35 413 POTENTIAL.
SQ SEQUENCE 413 AA; 44004 MW; F04DC71044F9E50F CRC64;

Query Match 18.1%; Score 404; DB 2; Length 413;
Best Local Similarity 32.3%; Pred. No. 1.9e-12;
Matches 141; Conservative 31; Mismatches 110; Indels 154; Gaps 14;

QY 1 AVENKEETPETDSEEVTIKANLIFANGSTQAEFGKTFEKATSEAYAYADTLKKDN 60
Db 30 AVNAEESTVSPVTATDVT-----TSKEALAIINKLSEDN 66
QY 61 GETVDVADKGYTLNKFAGKEKTPPEKPEVITKANL-IYADGKTQTAEFGKTFEEATA 119
Db 67 LN-NLDI-----QEVLAKEAGDILASDSADT--IKALLAEVTA 101
QY 120 EAVYADALAKKNGEYTVVDVADKGYTLNKFAGKEKTPPEKPEVITKANLIYADGKTQT 179
Db 102 EVTRL-----NEKMARDVADKRAIAADAA-----AFSELKD 132
QY 180 AEFKGTPEETAAYRAYDL-LAKENGKTYVDVADKGYTLNKFAGKEKTPPEKPEVIT 238
Db 133 AOLKAYEDLAKLAADTDLDDVAKIINDYTTKVENAKTAEDVK-----KIFESQNEVT- 186
QY 239 KANLIYADGKTQTAEFGKTFEATAAYRAYADLLAKENGKTYADLDEGGYTIINIRFAGKK 298
Db 187 -----RIKTERKALKAAALAKAKADAIE-----ILKKYG-----IGDYVI----- 220
```

```
QY 299 VDEKPEPMDTYKLIINGKTLKGETTEAADAATAEAKVFKQYANDNGVDGWTYDDATKT 358
Db 221 -----KLINNGKTAEGVT-----ALKD 237
QY 359 FVTEKPEVIDASELTPAVTTYKLIVNGKTLKGETTTKAVDAETAFAKQYANDNGVDG 418
Db 238 EILASKPAVIDADELPALPTYKLVINGKTLKGETTTKAVDAETAFAKQYANENGVDG 297
QY 419 VMTYDDATKFTVTTEM 434
Db 298 VMTYDDATKFTVTTEM 313

RESULT 9
Q53337
ID Q53337 PRELIMINARY; PRT; 60 AA.
AC Q53337;
DT 01-NOV-1996 (TREMBLrel. 01, Created)
DT 01-NOV-1996 (TREMBLrel. 01, Last sequence update)
DE PROTEIN G IGG FC BINDING DOMAIN.
OS Streptococcus sp.
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Streptococcaceae;
OC Streptococcus.
OX NCBI_TaxID=1306;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=93277681; PubMed=8503988;
RA Cai S.Y., Wang Y.Y., Yao Z.J.;
RT "Structure analysis of streptococcal protein G Fc binding domain.";
RL Sci. China B 36:75-80(1993).
DR EMBL; S62801; AAB27024.1; -.
DR HSSP; P06654; LPGX.
DR InterPro; IPR000724; IgG_bind_B.
DR Pfam; PF01378; IgG_binding_B; 1.
SQ SEQUENCE 60 AA; 6655 MW; 924567E0D6B513DB CRC64;

Query Match 13.2%; Score 294; DB 2; Length 60;
Best Local Similarity 96.8%; Pred. No. 4.2e-08;
Matches 56; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 377 VTTYKLIVNGKTLKGETTTKAVDAETAFAKQYANDNGVDGWYDDATKFTVTTEM 434
Db 1 MTTYKLIVNGKTLKGETTTKAVDAETAFAKQYANDNGVDGWYDDATKFTVTTEM 58

RESULT 10
Q9U8G8
ID Q9U8G8 PRELIMINARY; PRT; 3198 AA.
AC Q9U8G8;
DT 01-MAY-2000 (TREMBLrel. 13, Created)
DT 01-MAY-2000 (TREMBLrel. 13, Last sequence update)
DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)
DE LACUNIN PRECURSOR.
OS Manduca sexta (Tobacco hawkmoth) (Tobacco hornworm).
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Insecta;
OC Pterygota; Neoptera; Endopterygota; Lepidoptera; Glossata; Ditrysia;
OC Sphingioidea; Sphingidae; Sphinginae; Manduca.
OX NCBI_TaxID=7130;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=99457716; PubMed=10528409;
RA Nardi J.B., Martos R., Walden K.K., Lampe D.J., Robertson H.M.;
RT "Expression of lacunin, a large multidomain extracellular matrix
protein, accompanies morphogenesis of epithelial monolayers in Manduca
sexta.";
RL Insect Biochem. Mol. Biol. 29:883-897(1999).
DR EMBL; AF078161; AAF04457.1; -.
DR HSSP; P12111; 2KNT.
DR InterPro; IPR004094; Antistasin.
DR InterPro; IPR003598; Ig_G2.
DR InterPro; IPR003006; Ig_MHC.
```

```
DR InterPro: IPR002223; Kunitz_BPTI.
DR InterPro: IPR000884; TSP1.
DR InterPro: IPR002221; WAP.
DR Pfam: PF02822; Antistasin; 4.
DR Pfam: PF00047; ig; 2.
DR Pfam: PF00014; Kunitz_BPTI; 9.
DR Pfam: PF00095; wap; 1.
DR PRINTS: PR00759; BASICPTASE.
DR SMART: SM00408; IGC2; 2.
DR SMART: SM00131; KU; 10.
DR SMART: SM00209; TSP1; 7.
DR SMART: SM00217; WAP; 1.
DR PROSITE: PS00317; 4_DISULFIDE_CORE; 1.
DR PROSITE: PS00280; BPTI_KUNITZ_1; 8.
DR PROSITE: PS0279; BPTI_KUNITZ_2; 10.
DR PROSITE: PS0092; TSP1; 1.
KW Immunoglobulin domain; Serine protease inhibitor; Signal.
FT SIGNAL 1 21 POTENTIAL.
SQ SEQUENCE 3198 AA; 349364 MW; AB4ACD459CD09134 CRC64;

Query Match      8.2%; Score 182.5; DB 5; Length 3198;
Best Local Similarity 23.4%; Pred. No. 0.97;
Matches 106; Conservative 71; Mismatches 197; Indels 79; Gaps 21;

QY 11 TPTDSEEEVITKANLIFANGSTQTAFFKGFTEKAT-SEAYAYADTLKKDNGEYTVDVAD 69
Db 1028 TEESVAEEETK-----TTTEEVSGTSSESINSKTTMTTSLSDTGKTSV---- 1075

QY 70 KGYTLNFKAGKEKTP-EPEKEVITKANLIYADGKTQTAFFKGFTEKATAEAYADAL 128
Db 1076 -SEIETEMTVTSETSETPTGTSKDTMTSVSETESS---SVTEETTTTTVVENAT 1131

QY 129 KKDNGEYTVDVADKGYTLNFKAGKEKTPPEPEKEVITKANLIYADGKTQTAFFKGTPEE 188
Db 1132 DISSTEVT--ASDK-TWTMTSESEKTEETATETIV-----TKEVTSST 1176

QY 189 ATAERYADLLAKENGKYTVDVADKGYTLNFKAGKEKTPPE-----PKEEVTIKANLI 243
Db 1177 ETATSKTITLSEETGK--TSVAEESTTEKVTETVTTMTPEETGKTTITSEITIKTV- 1233

QY 244 YADGKTQTAFFKGFTEKATAEAYADLLAKENGKYTADEGGYVINIRFAGKVKDEKP 303
Db 1234 -----TEEPTDVG-S-EAITSKDTTVSTVASEETGKYSVSEEE---TVKTVVAEASPEPS 1284

QY 304 EEPMDYTKLILN-----GK-----TLKGETTTEAVDAATAEKFVKYANDNGVDGEW 350
Db 1285 TEAITSKDTKMTSITSEETGKTSVSEITVKTVTTEESTEPSSEITSEKTSVIS----- 1339

QY 351 TYDDATKFTT---VTEKPEVID-ASELTPAVT--TVKLINGKTLKGETTTKA---VDAE 401
Db 1340 TSEETGKTSVSEEVTVKTTVTDEATEITSTVSTETKTSVSGST--EELSTQASSKIESP 1397

QY 402 TAEKAFQYANDNGVDGVVTVYDDATKFTFVTEM 434
Db 1398 TTESGITSHTTESEESTVSTTEKGEVTSSETTEL 1430

RESULT 11
Q9LF88
ID Q9LF88 PRELIMINARY; PRT: 479 AA.
AC Q9LF88;
DT 01-OCT-2000 (TrEMBLrel. 15, Created)
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE LATE EMBRYOGENESIS ABUNDANT PROTEIN-LIKE.
GN F8J2.210.
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC eurosids II; Brassicales; Brassicaceae; Arabidopsis.
OX NCBI_TaxID=3702;
RN [1]
RP SEQUENCE FROM N.A.

RA Nyakatura G., Fartmann B., Dauner D., Sterr W., Holland R.,
RA Weichselgartner M., Mewes H.W., Rudd S., Lemcke K., Mayer K.F.X.,
RA Quetier F., Salanoubat M.;
RL Submitted (NOV-1999) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RA EU Arabidopsis sequencing project;
RL Submitted (AUG-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL: AL132969; CAB86908.1; -
DR InterPro: IPR004238; LEA.
DR Pfam: PF02987; LEA; 5.
SQ SEQUENCE 479 AA; 52084 MW; C39E000D910E2385 CRC64;

Query Match      8.1%; Score 182; DB 10; Length 479;
Best Local Similarity 24.1%; Pred. No. 0.11;
Matches 110; Conservative 59; Mismatches 195; Indels 92; Gaps 18;

QY 3 ENKETPTPEPDSEEEVITKANLIFANGSTQTAFFKGFTEKATSEAYAY-----ADTLKK 58
Db 75 DTAESTREGADIASEKAAGR-----DTTGEVRDSTAQKTKETADYTADKAREAKD 125

QY 59 DNGEYTVDVADKGYTLNFKAGKEKTPPEPEKEVITKANLIYADGKTQTAFFKGTFEAT 118
Db 126 KTADTKETAD--YAAEKAREAKDRTADTKETAEYTAEKA-REAKDTADKLGEYKDYT 182

QY 119 AEAYRYA-DALKKONGEYTVDVADKGYTLNFKAGKEKTPPEPEKEVITKANLIYADGKT 177
Db 183 AEKAEKADTTAEKLGEY-----KDYTVDRKAKEAKDTAEKAKETAEYTSDKA-RETKD 235

QY 178 QTAEPKGFTEKATAE-----AYRYADLLAKENGKYTVDVADKGYTLNFKAGKEKTP 229
Db 236 KTAEVKGEYKDYTAEKAKETADKAREAKDKTAEVKGEYRDYTAEK--ATETKDAGVSKIG 293

QY 230 EEPKEEV-TIKANLIYADGKTQ-----TAEKGTAEATAEAYRYADLLAKENGKYTADL 283
Db 294 ELKDSAVTAKRAMGFLSGKTEETKQKAVETKDTAKEKMDAGEAEARRKME----- 345

QY 284 EDGGYVINIRFAGKVKDEKPEPEPEPEPEPEPEPEPEPEPEPEPEPEPEPEPEPEPEPE 343
Db 346 -----MRLEGGKLDEDSR-----KTOOSTES--AADKAHETKDSVAQ 381

QY 344 NGVDGEYTVYDDATKFTTTEKPEVIDASELTPAVTVYKLINGKTLKGETTTKAVIDA--- 400
Db 382 RGEKGSIMGALGNMTGAIKSKLTGAT--TPSDEETRASAHGDESTGKTVV-AVDVKDT 438

QY 401 -----ETAEKAFQYANDNGVDGVVTVYDDATK 427
Db 439 RPYVATVILKEADQMTGTFNDVG-----EIDDEEK 469

RESULT 12
Q9AER7
ID Q9AER7 PRELIMINARY; PRT: 2402 AA.
AC Q9AER7;
DT 01-JUN-2001 (TrEMBLrel. 17, Created)
DT 01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE BAP-LIKE PROTEIN BHP.
OS Staphylococcus epidermidis.
OC Bacteria; Firmicutes; Bacillus/Clostridium group;
OC Bacillus/Staphylococcus group; Staphylococcus.
OX NCBI_TaxID=1282;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-RP62A;
RA Tormo M.A., Cuadrella C., Amorena B., Lasa I., Penades J.R.;
RT "The Bap Homolog Protein of Staphylococcus epidermidis RP62A Promotes
RT Biofilm Formation.";
RL Submitted (MAR-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL: AY028618; AAK29746.1; -
DR InterPro: IPR001298; Filamin.
DR InterPro: IPR001899; Gram_pos_anchor.
DR InterPro: IPR003410; HYR.
```

RP SEQUENCE FROM N.A.
RC STRAIN=CV. COLUMBIA;
RX MEDLINE=20083487; PubMed=10617197;
RA Lin X., Kaul S., Rounsley S.D., Shea T.P., Benito M.-I., Town C.D.,
RA Fujii C.Y., Mason T.M., Bowman C.L., Barnstead M.E., Feldblum T.V.,
RA Buell C.R., Ketchum K.A., Lee J.J., Ronning C.M., Koo H., Moffat K.S.,
RA Cronin L.A., Shen M., VanAken S.E., Umayam L., Tallon L.J., Gill J.E.,
RA Adams M.D., Carrera A.J., Creasy T.H., Goodman H.M., Somerville C.R.,
RA Copenhagen G.P., Preuss D., Nierman W.C., White O., Eisen J.A.,
RA Salzberg S.L., Fraser C.M., Venter J.C.;
RT -Sequence and analysis of chromosome 2 of the plant *Arabidopsis*
RT *thaliana*.";

```
DR InterPro: IPR003137; PA.
DR InterPro: IPR000209; Peptidase_s8.
DR Pfam: PF02225; PA; 1.
DR Pfam: PF00082; Peptidase_S8; 3.
DR PRINTS: PR00723; SUBTILISIN.
DR PROSITE: PS00588; FLAGELLA_BB_ROD; UNKNOWN_1.
DR PROSITE: PS00136; SUBTILASE_ASP; UNKNOWN_1.
DR PROSITE: PS00137; SUBTILASE_HIS; 1.
DR PROSITE: PS00138; SUBTILASE_SER; 1.
DR PROSITE: PS00139; SUBTILASE_SER; 1.
SQ SEQUENCE 1849 AA; 199614 MW; 219F0D44B15A091F CRC64;

Query Match
Best Local Similarity 7.6%; Score 170.5; DB 2; Length 1849;
Matches 117; Conservative 44; Mismatches 165; Indels 125; Gaps 25;

QY 73 TLNKFAGKEKTEPEEKKEEVTIKANLIYADGKTQTAEFKGTFEEATAEA---YYADALKK 130
DB 1415 TLNL---DSENTVTYNNKDKFTI-SGTISDDYKFYDLSINGNDVETSWSAVDYHSKEGIRK 1470
QY 131 DNGEYTVDVADKGYTLNLIK-----FAGKEKTEPEEK---VTIKANLIYA 173
DB 1471 -NPKHEVDLKKGNFTNVKVTIDIQGNSSQALVYVEPAKTLAEPVSKLLTKTANLQLL 1529
QY 174 DGKTQTAEFK-----GTFEATAEAYRYADLLAKENGKYTVDVADKGY-----TL 218
DB 1530 KATTDSEAKVWYSLDNGKTFNDVPADGFK-----VTENGTVQFKAVDK-YGNESKVKS 1583
QY 219 NIKFAGKEKTEPEEKKEEVTIKANL---IYADGKTQ---TAEFGTFAEATAEAYRYADL 271
DB 1584 EIKGLNKENQSPEDKELAKEKLNQAVDAGEKKDLDKYTADSKDFNDALKKA---KDV 1640
QY 272 LAKENGKYTADLEDGGYTIINIRFAGKKVDEKPEEPMYTKLIL-NGKTLKGTTTAEVDA 330
DB 1641 LADKNK-LADLQDAKALD--RAEQALTEKPAEP--TIPLLQGNNAVSNINTSSDNQV 1695
QY 331 ATAEEKVQYANDN-----GVDGEWY-----DDATKFTT-----VTEKP--- 365
DB 1696 AAPVHAEKDTKDNKNTTEGKDTVMFKSVLYTKDLKTRTAQAYSSLKLVTEKGLK 1755
QY 366 -----EVIDASELTPA--VTYIKLVI-----NGKTL-----KGETTITKAV 398
DB 1756 VYTFKGHYFYKVDNRNAYVRNVGTGKATLKRNSFYVOSNGKKASRKLKLGKGTITVYG 1815
QY 399 DAETAEKAFKQYA-----NDNGVD 417
DB 1816 DQYKALKHYKKYAIRIGEGRYIKSVNVRVD 1846

RESULT 15
Q93TY6 PRELIMINARY; PRT; 2276 AA.
AC Q93TY6;
DT 01-DEC-2001 (TrEMBLrel. 19, Created)
DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE BIOFILM-ASSOCIATED SURFACE PROTEIN.
GN BAP.
OS Staphylococcus aureus.
OC Bacteria, Firmicutes; Bacillus/Clostridium group;
OC Bacillus/Staphylococcus group; Staphylococcus.
OX NCBI_TaxID=1280;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-V329;
RX MEDLINE=21189316; PubMed=11292810;
RA Cucarella C.; Solano C.; Valle J.; Amorena B.; Lasa I.;
RA Penades J.R.;
RT "Bap, a Staphylococcus aureus Surface Protein Involved in Biofilm
RT Formation."
RL J. Bacteriol. 183:2888-2896(2001).
DR EMBL: AF288402; AAK38834.1; -.
SQ SEQUENCE 2276 AA; 238792 MW; 8623C60472CB0C0F CRC64;
```

```
Query Match
Best Local Similarity 7.6%; Score 170; DB 2; Length 2276;
Matches 110; Conservative 47; Mismatches 197; Indels 150; Gaps 25;

QY 8 TPETETDSEEEVTIKANLIFANGSTQTAEFKGTTEKATSEAYAYADTLKKNGEYTVDV 67
DB 1558 TSDATQVGTGAEPNSTVKLTFFDGTAT---GT-----AD-----DQGNVTIDI 1598
QY 68 ADK-----GYTLNIFAGKEKTEPEEK-----EVTIKANLIYADGKTQT 107
DB 1599 PANGDLNGEELQVATDKDGNTEPSSANVTDTTASDAPTVDVNTSDATQVGTGAEPNS 1658
QY 108 ABEFKGTFEATAEAYRYADALKKNGEYTVDVADK-----GYTLNIFAGKEKTEPEEK- 161
DB 1659 T-VKLTFFPDGTT-----ATGTADDQGNVTIDIPSNVDLNGEELQVATDKDGNTEPSS 1712
QY 162 -----BEVTIKANLIYADGKTQTAEFKGTTEEATAEAYRYADLLAKENGKY 207
DB 1713 ANVTDTTAPDAPTVDVNTSDATQVGTGAEPNST-VKLTFFPDGTT-----ATGTADDQGN 1766
QY 208 TVDVADK-----GYTLNIFAGKEKTEPEEK-----EVTIKANLIYADG 247
DB 1767 TIDIPSNVDLNGEELQVATDKDGNTEPSEKLTNVDTTASDAPTVDVNTSDASQVGTGA 1826
QY 248 KTQTAEFKGTFAEATAEAYRYADLLAKENGKYTABLE-----DGGYTIINIRFAGKKVDEK 302
DB 1827 EPNST-VKLTFFPDGTT-----ATGTADDQGNVTIDIPSNVDLNGEELQVATDK- DGN 1878
QY 303 PEEPMDTYKLIINGKTLKGTTTAEVDAATAEKVFKQYANDNGVDGEWYDDATKFTT 362
DB 1879 TSEPSA-----NVTDTTAPDAPTVDV-----TSDATQ---VT 1909
QY 363 EKPEVIDASELT-PAVTT-----YKLVI--NGKTLKGE---TTTKAVDAETA 406
DB 1910 GQAEPNSTVKLTFFPDGTTATGTADDQGNVTIDIPSNVDLNGEELQVATDKDGNTEPSS 1969
QY 407 FKQYANDNGVDGWYTVDDATKFTT 430
DB 1970 SANVTDTTAPDAP-TVNDVTSAT 1992

Search completed: October 29, 2002, 09:31:00
Job time : 35.364 secs
```


Result No.	Score	Query Match	Length	DB	ID	Description
1	1565	100.0	455	2	Q53291	Q53291 streptococc
2	1561	99.7	719	2	Q51912	Q51912 peptostrept
3	1226.5	78.4	992	2	Q51918	Q51918 peptostrept
4	167	10.7	664	2	Q53975	Q53975 streptococc
5	166	10.6	479	10	Q9LPR8	Q9LPR8 arabisidopsi
6	166	10.6	669	2	Q9SEW8	Q9SEW8 streptococc
7	151	9.6	448	10	Q9SKP0	Q9SKP0 arabisidopsi
8	149	9.5	448	10	Q96246	Q96246 arabisidopsi
9	148.5	9.5	2276	2	Q93TV6	Q93TV6 staphylococc
10	141.5	9.0	1110	13	Q91255	Q91255 petromyzon
11	141	9.0	1849	2	Q9S4K2	Q9S4K2 lactobacill
12	137.5	8.8	992	16	Q9WZA6	Q9WZA6 thermotoga
13	137.5	8.8	5458	5	Q9U459	Q9U459 plasmodium
14	136	8.7	506	2	Q53837	Q53837 salmonella
15	134.5	8.5	3198	5	Q9U8G8	Q9U8G8 manduca sex
16	132.5	8.5	737	16	Q9JYK4	Q9JYK4 neisseria m

```
QY 121 AYRYADALKKDNGETYVDVADKGYTLNKFAGKEKTPPEPKKEEVTIKANLIYADGKTQTA 180
|||||
Db 142 AYRYADALKKDNGETYVDVADKGYTLNKFAGKEKTPPEPKKEEVTIKANLIYADGKTQTA 201
|||||
QY 181 EFKGTFEEATAEAYRYADLLAKENGYTVVDVADKGYTLNKFAGKEKTPPEPKKEEVTIKA 240
|||||
Db 202 EFKGTFEEATAEAYRYADLLAKENGYTVVDVADKGYTLNKFAGKEKTPPEPKKEEVTIKA 261
|||||
QY 241 NLIYADGKTQTAEFKGFPAEATAEAYRYADLLAKENGYTADLEDGGYTTINIRFAGKKVD 300
|||||
Db 262 NLIYADGKTQTAEFKGFPAEATAEAYRYADLLAKENGYTADLEDGGYTTINIRFAGKKVD 321
|||||
QY 301 EKPEE 305
|||||
Db 322 EKPEE 326
|||||

RESULT 2
Q51912
ID Q51912 PRELIMINARY; PRT; 719 AA.
AC Q51912;
DT 01-NOV-1996 (TrEMBLrel. 01, Created)
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DE PROTEIN L PRECURSOR.
OS Peptostreptococcus magnus.
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Clostridiaceae;
OC Finegoldia.
OX NCBI_TaxID=1260;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=312;
RX MEDLINE=90215984; PubMed=2108927;
RA Kastern W., Holst E., Nielsen E., Sjobring U., Bjorck L.;
RT "Protein L, a bacterial immunoglobulin-binding protein and possible
virulence determinant.";
RL Infect. Immun. 58:1217-1222(1990).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=312;
RX MEDLINE=92316971; PubMed=1618782;
RA Bjoerck L., Sjoerbring U., Kastern W.;
RT "Structure of peptostreptococcal protein L and identification of
repeated immunoglobulin light chain-binding domain.";
RL J. Biol. Chem. 267:12820-12825(1992).
DR EMBL: M86697; AAA25612.1; -.
DR InterPro: IPR003147; B1.
DR InterPro: IPR001899; Gram_pos_anchor.
DR Pfam: PF02246; B1; 5.
DR Pfam: PF00746; Gram_pos_anchor; 1.
KW SIGNAL.
FT SIGNAL.
FT CHAIN 1 18 POTENTIAL.
FT CHAIN 19 719 PROTEIN L.
SQ SEQUENCE 719 AA; 78983 MW; 963A8D76D5E34DD2 CRC64;

Query Match 99.7%; Score 1561; DB 2; Length 719;
Best Local Similarity 100.0%; Pred. No. 2.5e-78;
Matches 304; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 VENKEETPETDSEEEVTIKANLIFANGSTQTAEFKGFEEKATSEAYAYADTLKKDNG 61
|||||
Db 95 VENKEETPETDSEEEVTIKANLIFANGSTQTAEFKGFEEKATSEAYAYADTLKKDNG 154
|||||
QY 62 EYTVVDVADKGYTLNKFAGKEKTPPEPKKEEVTIKANLIYADGKTQTAEFKGFEEATAEA 121
|||||
Db 155 EYTVVDVADKGYTLNKFAGKEKTPPEPKKEEVTIKANLIYADGKTQTAEFKGFEEATAEA 214
|||||
QY 122 YRYADALKKDNGETYVDVADKGYTLNKFAGKEKTPPEPKKEEVTIKANLIYADGKTQTA 181
|||||
Db 215 YRYADALKKDNGETYVDVADKGYTLNKFAGKEKTPPEPKKEEVTIKANLIYADGKTQTA 274
|||||
QY 182 FKGTFFEEATAEAYRYADLLAKENGYTVVDVADKGYTLNKFAGKEKTPPEPKKEEVTIKAN 241
|||||
```

```
Db 275 FKGTFFEEATAEAYRYADLLAKENGYTVVDVADKGYTLNKFAGKEKTPPEPKKEEVTIKAN 334
|||||
QY 242 LIYADGKTQTAEFKGFPAEATAEAYRYADLLAKENGYTADLEDGGYTTINIRFAGKKVDE 301
|||||
Db 335 LIYADGKTQTAEFKGFPAEATAEAYRYADLLAKENGYTADLEDGGYTTINIRFAGKKVDE 394
|||||
QY 302 KPEE 305
|||||
Db 395 KPEE 398
|||||

RESULT 3
Q51918
ID Q51918 PRELIMINARY; PRT; 992 AA.
AC Q51918;
DT 01-NOV-1996 (TrEMBLrel. 01, Created)
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DE PROTEIN L PRECURSOR.
OS Peptostreptococcus magnus.
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Clostridiaceae;
OC Finegoldia.
OX NCBI_TaxID=1260;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=3316;
RX MEDLINE=95078460; PubMed=7987012;
RA Murphy J.P., Trowern A.R., Duggieby C.J.;
RT "Nucleotide sequence of the gene for peptostreptococcal protein L.";
RL DNA Seq. 4:259-265(1994).
DR EMBL: L04466; AAA67503.1; -.
DR HSP: Q51911; LGAB.
DR InterPro: IPR003147; B1.
DR InterPro: IPR002988; GA.
DR InterPro: IPR001899; Gram_pos_anchor.
DR Pfam: PF02246; B1; 4.
DR Pfam: PF01468; GA; 4.
DR Pfam: PF00746; Gram_pos_anchor; 1.
KW SIGNAL.
FT SIGNAL.
FT CHAIN 1 24 POTENTIAL.
FT CHAIN 25 992 PROTEIN L.
SQ SEQUENCE 992 AA; 108700 MW; 9CFF5771578A5DCE CRC64;

Query Match 78.4%; Score 1226.5; DB 2; Length 992;
Best Local Similarity 79.9%; Pred. No. 8.5e-60;
Matches 246; Conservative 22; Mismatches 31; Indels 9; Gaps 4;

QY 2 VENKEETPETDSEEEVTIKANLIFANGSTQTAEFKGFEEKATSEAYAYADTLKKDNG 61
|||||
Db 241 MERKLEKETPE--PEEEVTIKANLIFADGSGTQNAEFKGFKAQVSDAYAYADALKKNG 298
|||||
QY 62 EYTVVDVADKGYTLNKFAGKEKTPPEPKKEEVTIKANLIYADGKTQTAEFKGFEEATAEA 121
|||||
Db 299 EYTVVDVADKGYTLNKFAGKEKTPPEPKKEEVTIKANLIFADGSGTQNAEFKGFEEATAEA 358
|||||
QY 122 YRYADALKKDNGETYVDVADKGYTLNKFAGKEKTPPEPKKEEVTIKANLIYADGKTQTA 179
|||||
Db 359 YAYADLLAKENGYTADLEDGGYTTINIRFAGKEKTPPEPKKEEVTIKANLIYADGKIQT 418
|||||
QY 180 AEPKGFEEATAEAYRYADLLAKENGYTVVDVADKGYTLNKFAGKEKTPPEPKKEEVT 237
|||||
Db 419 AEPKGFEEATAEAYRYADLLAKENGYTADLEDGGYTTINIRFAGKEKTPPEPKKEEVT 478
|||||
QY 238 IKANLIYADGKTQTAEFKGFPAEATAEAYRYADLLAKENGYTADLEDGGYTTINIRFAGK 297
|||||
Db 479 IKVNLIFADGKTQTAEFKGFPAEATAEAYRYADLLAKENGYTADLEDGGYTTINIRFAGK 538
|||||
QY 298 KVDEKPEE 305
|||||
Db 539 ---EQPGE 543
|||||
```

```
RESULT 4
Q53975 Q53975 PRELIMINARY; PRT; 664 AA.
AC Q53975;
DT 01-NOV-1996 (T-EMBLrel. 01, Created)
DT 01-NOV-1996 (T-EMBLrel. 01, Last sequence update)
DT 01-JUN-2001 (T-EMBLrel. 17, Last annotation update)
DE IMMUNOGLOBIN G BINDING PROTEIN MIG PRECURSOR (IGG BINDING PROTEIN
DE MIG).
GN MIG.
OS Streptococcus dysgalactiae.
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Streptococcaceae;
OC Streptococcus.
OX NCBI_TaxID=1334;
RN [1]
RP SEQUENCE FROM N.A., AND FUNCTION.
RC STRAIN=SC1;
RX MEDLINE=94192673; PubMed=8143736;
RA Jonsson H., Mueller H.-P.;
RT "The type-III Fc receptor from Streptococcus dysgalactiae is also an
RT alpha 2-macroglobulin receptor.";
RL Eur. J. Biochem. 220:819-826(1994).
RN [2]
RP FUNCTION.
RX MEDLINE=99081771; PubMed=9864244;
RA Vasi J., Svensson J., Frick I.-M., Mueller H.-P.;
RT "Five homologous repeats of the protein G-related protein MIG
RT cooperate in binding to goat immunoglobulin G.";
RL Infect. Immun. 67:413-416(1999).
CC -1- FUNCTION: BINDS TO THE CONSTANT FC REGION OF IGG WITH HIGH
CC AFFINITY. N-TERMINUS MEDIATES BINDING TO PLASMA PROTEINASE
CC INHIBITOR ALPHA 2-MACROGLOBULIN AFTER COMPLEX FORMATION WITH
CC PROTEASES.
CC -1- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN. CELL WALL (BY
CC SIMILARITY).
CC -1- SIMILARITY: TO OTHER STREPTOCOCCAL AND STAPHYLOCOCCAL PROTEINS IN
CC THE REGION OF THE MEMBRANE ANCHOR. C-TERMINAL IS SIMILAR TO OTHER
CC STREPTOCOCCAL TYPE-III FC RECEPTORS.
DR EMBL: Z29666; CAA82764.1;
DR HSP: P06654; IPGX
DR InterPro: IPR001899; Gram_pos_anchor.
DR InterPro: IPR000724; Igg_bind_B.
DR InterPro: IPR000531; TonB_boxC.
DR Pfam: PF00746; Gram_pos_anchor; 1.
DR Pfam: PF01378; Igg_binding_B; 5.
DR PRINTS: PR00015; GP0SANCHOR.
DR PROSITE: PS00343; GRAM_POS_ANCHORING; UNKNOWN_1.
DR PROSITE: PS00430; TONB_DEPENDENT_REC_1; UNKNOWN_1.
DR Igg-binding protein; Repeat; Transmembrane; Cell wall; Signal.
KW SIGNAL 1 30
FT CHAIN 31 664
FT DOMAIN 31 208
FT DOMAIN 559 638
FT DOMAIN 639 659
FT TRANSMEM 660 664
FT DOMAIN 664 664
FT DOMAIN 87 189
FT DOMAIN 87 110
FT REPEAT 166 189
FT REPEAT 209 558
FT DOMAIN 209 278
FT REPEAT 209 278
FT REPEAT 279 348
FT REPEAT 349 418
FT REPEAT 419 458
FT REPEAT 489 558
FT DOMAIN 592 621
FT DOMAIN 630 635
FT SEQUENCE 664 AA; 72023 MW; 4A4ICEEF7977862A CRC64;
Query Match 10.7%; Score 167; DB 2; Length 664;
Best Local Similarity 27.4%; Pred. No. 0.08;
```

```
Matches 84; Conservative 49; Mismatches 134; Indels 40; Gaps 16;
Qy 7 ETPETPTDSEEVTKANLIFANGSTQTAEFKGTAEKATSEAYAYADTLKKDNGEYTD 66
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
214 DAPELTPALTYKLVVGN-TFSGETTTKADTATAKEFKQ-YATANNV---DGEWSYD 268
Qy 67 VADKGYTLNKA---GKEKTPPEPKKEVITKANLIYADGKTQTAEFKGTAEKATSEAY 123
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
269 DATKTFVTEKPAVIDAPELTPALTYKLVKGNF--SGETTT--KAVIDAEKAEKAFK 323
Qy 124 -YADALKKDNGETYVDVADKGYTLNKA---GKEKTPPEPKKEVITKANLIYADGKTQ 179
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
324 QYATANNVD--GEWSYDDATKTFVTEKPAVIDAPELTPALTYKLVKGNF--SGETTT 380
Qy 180 AEFKGTAEKATSEAYRYADLLAKENGKTYVDVADKGYTLNKA---GKEKTPPEPKKE 236
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
381 KADAATAEKEFKOYATANGV---DGEWSYDDATKTFVTEKPAVIDAPELTPALTYK 437
Qy 237 TIKANLIYADGKTQTAEFKGTAEKATSEAYRYADLLAKENGKTYADLEDGTYTINIRFAG 296
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
438 IVKGNF--SGETTT--KAVIDAEKAEKAFKQ---YANENGKTY-----GEWSYDDATKT 483
Qy 297 KKVDEKP 303
Db : : : :
484 FTVTEKP 490
RESULT 5
Q9LF88 Q9LF88 PRELIMINARY; PRT; 479 AA.
AC Q9LF88;
DT 01-OCT-2000 (T-EMBLrel. 15, Created)
DT 01-OCT-2000 (T-EMBLrel. 15, Last sequence update)
DT 01-DEC-2001 (T-EMBLrel. 19, Last annotation update)
DE LATE EMBRYOGENESIS ABUNDANT PROTEIN-LIKE.
GN F8J2.210.
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; Core eudicots; Rosidae;
OC eurosids II; Brassicales; Brassicaceae; Arabidopsids.
OX NCBI_TaxID=3702;
RN [1]
RP SEQUENCE FROM N.A.
RA Nyakatura G., Partmann B., Dauner D., Sterr W., Holland R.,
RA Weichselgartner M., Mewes H.W., Rudd S., Lemcke K., Mayer K.F.X.,
RA Quettier F., Salanoubat M.;
RL Submitted (NOV-1999) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RA EU Arabidopsids sequencing project;
RL Submitted (AUG-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL: AL132969; CAB86908.1;
DR InterPro: IPR004238; LEA.
DR Pfam: PF02987; LEA; 5.
SQ SEQUENCE 479 AA; 52084 MW; C39E00D910E2385 CRC64;
Query Match 10.6%; Score 166; DB 10; Length 479;
Best Local Similarity 25.8%; Pred. No. 0.062;
Matches 82; Conservative 43; Mismatches 137; Indels 56; Gaps 12;
Qy 3 ENKETPTPTDSEEVTKANLIFANGSTQTAEFKGTAEKATSEAYAY----ADTLKK 58
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
75 DTAESTREGADIASEKAAGMR-----DTTGEVDRDSTAQKTETADYADKAREAKD 125
Qy 59 DNGEYTVDVADKGYTLNKAIFAGKKTPEEPKEEVTIKANLIYADGKTQTAEFKGTAE 118
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
126 KTADKTETAD--YAAEKAREAKDRTADKTETAEYAEKA-REAKOKTADKLGKDYDT 182
Qy 119 AEAYRYA-DALKKDNGETYVDVADKGYTLNKAIFAGKKTPEEPKEEVTIKANLIYADGKT 177
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
183 AEKAEAKDPTAEKLGKY-----KDYTVDKAKEAKDKTAEKATSEAYTSKKA-RETKD 235
Qy 178 QTAEFKGTAEKATSEAY-----AYRYADLLAKENGKTYVDVADKGYTLNKAIFAGK 229
```

```

Db 236 KTAEKVGEYDYTAERAKETADKAREAKOTAEKVGEYDYTAEK--ATETKDAGVSKIG 293
QY 230 EEPKEEV-TIKANLIYADGKTO-----TAEFGTFAEATAEAYRYADLLAKENGKYTADL 283
Db 294 ELKDSAVDTAKRAMGFLUSGTEETKOKAVETKDTAKEKMDKEAGEEARRKKEE----- 345
QY 284 EDGGYTINIRFAGKKVDE 301
Db 346 -----MRLEGGKKLDE 355

RESULT 6
Q93EM8 PRELIMINARY; PRT; 669 AA.
ID Q93EM8
AC Q93EM8;
DT 01-DEC-2001 (TrEMBLrel. 19, Created)
DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DE 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE MIG PRECURSOR.
OS Streptococcus dysgalactiae.
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Streptococcaceae;
OC Streptococcus.
OX NCBI_TaxID=1334;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ATCC 43078;
RX MEDLINE=21437603; PubMed=11553540;
RA Song X.M., Perez-Casal J., Bolton A., Potter A.A.;
RT "Surface-Expressed Mig Protein Protects Streptococcus dysgalactiae
RT against Phagocytosis by Bovine Neutrophils.";
RL Infect. Immun. 69:6030-6037(2001).
DR EMBL; AF354651; AAL09476.1; -.
KW Signal.
FT SIGNAL.
FT CHAIN
FT CHAIN
SQ SEQUENCE 669 AA; 72682 MW; 5C8982B952029B03 CRC64;

Query Match 10.6%; Score 166; DB 2; Length 669;
Best Local Similarity 27.5%; Pred. No. 0.091;
Matches 85; Conservative 48; Mismatches 132; Indels 44; Gaps 16;

QY 7 ETPTPTDSEEVTIKANLIFANGSTQTAEFKGTPEKATSEAYAYADTLKONGEYTV 66
Db 214 DAPELTPALTYKLIVKGN-TFSGETTKAIDTATAEKFQ-YATANNV---DGEWSYD 268
QY 67 VADGKTYLNIKFA--GKEKTPPEKPEEVIKANLIYADGKTQTAEFKGTPEEATAEA-- 121
Db 269 DATKFTFTVEKPAVIDALELTPALTYKLIVKGNTE--SGETTTK----AIDAATAEKEF 322
QY 122 YRYADALKDNGEYTVADVADKGYTLNIKFA--GKEKTPPEKPEEVIKANLIYADGKTQ 178
Db 323 KOYATANNVD-GEWSYDYATKFTFTVEKPAVIDAPELTPALTYKLIVKGNTE--SGETT 379
QY 179 TAEFGTFAEATAEAYRYADLLAKENGKYTVADVADKGYTLNIKFA--GKEKTPPEKPEE 235
Db 380 TKAIIDAATAEKEFKOYATAN---NVDEWSYDDATKFTFTVEKPAVIDAPELTPALTYK 436
QY 236 VTIKANLIYADGKTQTAEFKGTFAEATAEAYR-YADLLAKENGKYTADLGGYTINIRF 294
Db 437 LIVKGNFTF--SGETTT---KAVDAETAERAKFOYA-----TANNVDGEWSYDDAT 481
QY 295 AGKKVDEKP 303
Db 482 KFTVTEKP 490

RESULT 7
Q9SKP0 PRELIMINARY; PRT; 448 AA.
ID Q9SKP0
AC Q9SKP0;
DT 01-MAY-2000 (TrEMBLrel. 13, Created)
DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
```

```

DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE LATE EMBRYOGENESIS ABUNDANT PROTEIN (ATECP63).
GN AT2G36640.
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC eurosids II; Brassicales; Brassicaceae; Arabidopsids.
OX NCBI_TaxID=3702;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=CV. COLUMBIA;
RX MEDLINE=20083487; PubMed=10617197;
RA Lin X., Kaul S., Rounsley S.D., Shea T.P., Benito M.-I., Town C.D.,
RA Fujii C.Y., Mason T.M., Bowman C.L., Barnstead M.E., Feldblum T.V.,
RA Buell C.R., Ketchum K.A., Lee J.J., Ronning C.M., Koo H., Moffat K.S.,
RA Cronin L.A., Shen M., VanAken S.E., Umavam L., Tallon L.J., Gill J.E.,
RA Adams M.D., Carrera A.J., Creasy T.H., Goodman H.M., Somerville C.R.,
RA Copenhaver G.P., Preuss D., Nierman W.C., White O., Eissen J.A.,
RA Salzberg S.L., Fraser C.M., Venter J.C.;
RT "Sequence and analysis of chromosome 2 of the plant Arabidopsis
RT thaliana.";
RL Nature 402:761-768(1999).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=CV. COLUMBIA;
RX Submitted (MAR-2000) to the EMBL/GenBank/DBSJ databases.
RA Lin X.;
RL EMBL; AC006282; AAD20140.1; -.
DR InterPro; IPR004238; LEA.
DR Pfam; PF02987; LEA; 4.
SQ SEQUENCE 448 AA; 48492 MW; C72563D4194DD9BF CRC64;

Query Match 9.6%; Score 151; DB 10; Length 448;
Best Local Similarity 24.5%; Pred. No. 0.38;
Matches 79; Conservative 38; Mismatches 107; Indels 98; Gaps 14;

QY 3 ENKEETPTDSEEVTIKANLIFANGSTQTAEFKGTPEKATSEAYAYADTLKNDGE 62
Db 79 EAAESTKEGAQIASEKAVGAK-----DATVEKAKET----ADYTAEKVGE 119
QY 63 YIVDVADKGYTLNIRFAGKEKTPPEKPEEVIKANLIYADGKTQTAEFKGTPEEATAEAY 122
Db 120 Y-----KDYTVDKAKEAKDTTAEKAKETANYTADKA-VEAKDKTAEKIGEYK----- 165
QY 123 RVADALKDNGEYTVADVADKGYTLNIRFAGKEKTPPEKPEEVIKANLIYADGKTQTAEF 182
Db 166 -----DYAVDKA-----VEAKDKTAEKAKETANYTADKA-KEAKDKTAEK 204
QY 183 KGTFFEATAE-----AYRYADLLAKENGKYTVADVADKGYTLNIRFAGKEKTPPE 230
Db 205 VGEYKDYTVDKAVEARDYTAERAKAIAEAKDKTAEKTEY-----KDYTVKATEGKDVTS 258
QY 231 ---EPKEEV--TIKANLIYADGKTQTAEFKGTFAEATAEAYRYADLLAKENGKYTADL 285
Db 259 KIGELKDSAVEAKRAMGFLSGKTEEAKGKAVETKDT-----AKEN-----MEK 302
QY 286 GGVTI-----NIRFAGKKVDEK 302
Db 303 AGEVTRQKWEEMRLEGGKELKEE 324

RESULT 8
Q96246 PRELIMINARY; PRT; 448 AA.
ID Q96246
AC Q96246;
DT 01-FEB-1997 (TrEMBLrel. 02, Created)
DT 01-FEB-1997 (TrEMBLrel. 02, Last sequence update)
DE 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE LEA PROTEIN IN GROUP 3.
GN ATECP63.
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
```

OC eucosids II; Brassicales; Brassicaceae; Arabidopsis.
OX NCBI_TaxID=3702;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN=COLUMBIA; TISSUE=DRY SEED;

RX MEDLINE=97169149; PubMed=9016956;

RA Yang H., Saitou T., Komeda Y., Harada H., Kamada H.;

RT "Arabidopsis thaliana ECP63 encoding a LEA protein is located in

RT chromosome 4.";

RL Gene 184:83-88(1997).

DR EMBL; D64140; BAA1017.1; -.

DR InterPro; IPR004238; LEA.

DR Pfam; PF02987; LEA; 4.

SQ SEQUENCE 448 AA; 48534 MW; DC37099F18FBA85E CRC64;

Query Match 9.5%; Score 149; DB 10; Length 448;

Best Local Similarity 24.5%; Pred. No. 0.49;

Matches 79; Conservative 38; Mismatches 107; Indels 98; Gaps 14;

QY 3 ENKEETPETDSEEVITKANLIFANGSTQTAEPKGTPEKATSEAYAYADTLKDNGE 62

DB 79 EAESTKEGAIASEKAVGAK-----DATVEKAKET---ADYAEKVGE 119

QY 63 YTVDAVKGTYLTKFKAGKEKTPPEKPEVITKANLIYADGKTQTAEPKGTPEATAEAY 122

DB 120 Y-----KDYTVDAKAEKADTTAEKAKETANYTADKA-VEAKDKTAEKIGEK----- 165

QY 123 RYADALKDNGEYTVDAVKGTYLTKFKAGKEKTPPEKPEVITKANLIYADGKTQTAEP 182

DB 166 -----DYAVDKA-----VEAKDKTAEKAKETSNTADKA-KEAKDKTAEK 204

QY 183 KGTEPEATAE-----AYRYADLLAKENGKTYTVDAVKGTYLTKFKAGKEKTPPE 230

DB 205 VGEKDYTVDAKAVARDYTAEKAEAKDKTAEKIGY-----KDYVEKATEGKDYTVS 258

QY 231 ---EPKEEV---TIKANLIYADGKTQTAEPKGTPEATAEAYRYADLLAKENGKTYADLE 285

DB 259 KLGLKDSAVETAKRAMGFLSGKTEAKGAVETKDT-----AKEN-----MEK 302

QY 286 GGYTI-----NIRPAGKVDK 302

DB 303 AGEVTRQKMEEMRLGKELKEE 324

RESULT 9

Q93TY6

ID Q93TY6 PRELIMINARY; PRT; 2276 AA.

AC Q93TY6;

DT 01-DEC-2001 (Tremblrel. 19, Created)

DT 01-DEC-2001 (Tremblrel. 19, Last sequence update)

DT 01-DEC-2001 (Tremblrel. 19, Last annotation update)

DE BIOFILM-ASSOCIATED SURFACE PROTEIN.

GN BAP.

OS Staphylococcus aureus.

OC Bacteria; Firmicutes; Bacillus/Clostridium group;

OC Bacillus/Staphylococcus group; Staphylococcus.

OX NCBI_TaxID=1280;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN=V329;

RX MEDLINE=21189316; PubMed=11292810;

RA Cucarella C., Solano C., Valle J., Amorena B., Lasa I.,

RA Penades J.R.;

RT "Bap, a Staphylococcus aureus Surface Protein Involved in Biofilm

RT Formation.";

RL J. Bacteriol. 183:2888-2896(2001).

DR EMBL; AF288402; AAK38834.1; -.

SQ SEQUENCE 2276 AA; 238792 MW; 8623C60472CB0C0F CRC64;

Query Match

Best Local Similarity 9.5%; Score 148.5; DB 2; Length 2276;

Matches 72; Conservative 36; Mismatches 135; Indels 97; Gaps 15;

QY 8 TPETPTDSEEVITKANLIFANGSTQTAEPKGTPEKATSEAYAYADTLKKNGEYTVDV 67

DB 1644 TSDATQVTGQAEPNSTVKLFPDGTAT-----GT-----AD-----DQNYTIDI 1684

QY 68 ADK-----GYTLNKFAGKEKTPPEPK-----EVTIKANLIYADGKTQ 107

DB 1685 PSNVDLNGGELQVATDKDGTSEPSANVTDTTAPDAPTVNDVTSATQVTGQAEPS 1744

QY 108 AEFKGTPEEATAEAYRYADALKKONGEYTVDAVK-----GYTLNKFAGKEKTPPEPK- 161

DB 1745 T-VKLTFPDGTT-----ATGTADDOGNYTIDIPSNVNLNGGELQVATDKDGTSEPKL 1798

QY 162 -----EVTIKANLIYADGKTQTAEPKGTPEEATAEAYRYADLLAKENGKY 207

DB 1799 TNVTDTTASDAPTVNDVTSASQVTGQAEPNST-VKLTFPDGTT-----ATGTADDOGNY 1852

QY 208 TVDQVADK-----GYTLNKFAGKEKTPPEPK-----EVTIKANLIYADG 247

DB 1853 TIDIPSNVNLNGGELQVATDKDGTSEPSANVTDTTAPDAPTVNDVTSATQVTGQA 1912

QY 248 KTQTAEPKGTFAEATAEAYRYADLLAKENGKYTADLEDGG 287

DB 1913 EPNST-VKLTFPDGTT-----ATGTADDOGNYTIDIPSNVNLNGGELQVATDKDGTSEPKL 1946

RESULT 10

Q91255

ID Q91255 PRELIMINARY; PRT; 1110 AA.

AC Q91255;

DT 01-NOV-1996 (Tremblrel. 01, Created)

DT 01-NOV-1996 (Tremblrel. 01, Last sequence update)

DT 01-DEC-2001 (Tremblrel. 19, Last annotation update)

DE NF-180.

OS Petromyzon marinus (Sea lamprey).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Hyperoartia;

OC Petromyzontiformes; Petromyzontidae; Petromyzon.

OX NCBI_TaxID=7757;

RN [1]

RP SEQUENCE FROM N.A.

RC TISSUE=CNS;

RX MEDLINE=95287814; PubMed=7770000;

RA Jacobs A.J., Kamholz J., Selzer M.E.;

RT "The single lamprey neurofilament subunit (NF-180) lacks

RT multiphosphorylation repeats and is expressed selectively in

RT projection neurons.";

RL Brain Res. Mol. Brain Res. 29:43-52(1995).

DR EMBL; U19361; AAA80106.1; -.

DR InterPro; IPR001664; IF.

DR Pfam; PF00038; Tropomyosin.

DR PRINTS; PR00194; TROPOMYOSIN.

DR PROSITE; PS00226; IF; UNKNOWN_1.

SQ SEQUENCE 1110 AA; 123818 MW; 6558DA73DAF6974C CRC64;

Query Match

Best Local Similarity 9.0%; Score 141.5; DB 13; Length 1110;

Matches 87; Conservative 38; Mismatches 138; Indels 79; Gaps 13;

QY 1 AVENKEETPETPTDSEEVITKANLIFANGSTQTAEPKGTPEKATSEAYAYADTLK 59

DB 482 SAEKDEEEEEEEKEEEEE-----AAEAEEDRGRKEGEAAEAEAEVEEKEE 532

QY 60 NGEYTVDAVADGKYT-----LNLFAGKEKTPPEKPEVITKANLIYADGKT 105

DB 533 AEEAEVEEAE 584

QY 106 QTAEFKGTPEEATAE 156

DB 585 EAAAKAE 638

QY 157 PEE---PKEEVTIKANLIYADGKTQTAEPKGTPEEATAEAYRYADLLAKENGKYTVDAVK 214

DB 639 EEEAEAEAEAEVTSK-----KAKTQEAEE-----VEEEAEAEAEAEAEAEAE 684

Q9W2A6

SHAN YIN-FU¹/HN,
Shan Z.X., Yu X.B., Li X.R., Ma C.L., Fang J.M., Lu J.H., X J.;
"Molecular cloning and structure analysis of the Plasmodium falciparum

```
RT erythrocyte membrane-associated giant protein Ag332 (Pf332) gene." ;
RL Submitted (FEB-2001) to the EMBL/GenBank/DDBJ databases.
DR EMBL; AF20180; AAF15293.3; -
DR InterPro; IPR001313; PUM.
SQ SEQUENCE 5458 AA; 615269 MW; 4F6A37D92BD80172 CRC64;

Query Match      8.8%; Score 137.5; DB 5; Length 5458;
Best Local Similarity 20.3%; Pred. No. 37;
Matches 67; Conservative 55; Mismatches 138; Indels 67; Gaps 12;

QY   3 ENKET---PETPTDSEEEVTKANLIIFANGSTOTAEF-----KGTPEKATSEAYAVADT 55
Db    ||| : | | | | | | | | | | | | | | | | | | | | | | | | | | | |
1928 EKSESSAISVEVKKEESISEK--IIVEEVSVTTEIVROESTTEKIVKDVPEDF 1984
QY   56 LKK-----DNGETTVDVADGYTLNKFAGKEKTPEEPKEEVTIKANLIYAD 102
Db    :: : | | | | | | | | | | | | | | | | | | | | | | | | | |
1985 VEOTDSVTEKVIEEGSNTVAEDVEEK-----ESASDEHEGEDSVNAQV---- 2030
QY   103 GKQTAEFKTFEFATAEAARYADALKNDNGCEYTVDVADKGTYLNINIKFAGKEKTPPEPK 162
Db    ||| : | | | | | | | | | | | | | | | | | | | | | | | | |
2031 ----TCBKSKSYTKRIVDEVSR-TTEEINGSGVTEGVDETG-----SVTEEII 2074
QY   163 EVTIKANLIYADGKTQAETFGTPEETAARAYRADLLAKENGKYTVDVADKGTYLNINIKF 222
Db    ||| : | | | | | | | | | | | | | | | | | | | | | | | | |
2075 EANTVEEVVEDGSVTEEWEDGSVIQWEDGSTVEEIQQENGGSVTEIVEEGSVN--- 2131
QY   223 AGREKTEPEEPKEEVTIKANLIYADGKTQTAEFTFAATAEAAY-RYADLLAKENGKYTA 281
Db    ||||| : | | | | | | | | | | | | | | | | | | | | | | |
2132 -----EEVEEYSVGSEDVETRYVTEVEEBSVVVEEIEEGSVVEEIEBEGSVYE 2184
QY   282 DL-EDGGYTIN-LIRFAGKKVDKPEE 305
Db    :: | : | : | : | | | | | | | | | | | | | | | | | | | |
2185 EIVEEGSVVEEIVEEGSVVEEIVEE 2211

RESULT 14
ID Q53837 PRELIMINARY; PRT; 506 AA.
AC Q53837;
DT 01-NOV-1996 (TrEMBLrel. 01, Created)
DD 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DI 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE PHASE-2 FLAGELLIN STRUCTURAL PROTEIN.
OS FLJB.
GN Salmonella cholerae-suis (Salmonella enterica).
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OX Salmonella.
NCBI_TaxID=591;
[1]
SEQUENCE FROM N.A.
RC STRAIN=SEROTYPE CHOLERAESUIS, ATCC 7001;
RX MEDLINE=95325331; PubMed=7541401;
RA Vanegas R.A., Joys T.M.;
DE "Molecular analyses of the phase-2 antigen complex 1.2. . of Salmonella spp.";
RT RT
RL J. Bacteriol. 177:3863-3864(1995).
DL EMBL: U17175; AAC43352.1; -
DR InterPro; IPR001029; Flagellin_C.
DR InterPro; IPR001492; Flagellin_N.
DR Pfam; PF00700; Flagellin_C; 1.
DR Pfam; PF00669; Flagellin_N; 1.
DR PRINTS; PR00207; FLAGELLIN.
DR ProDom; PD000316; Flagellin_C; 1.
SQ SEQUENCE 506 AA; 52590 MW; B9E41977F0A96A20 CRC64;

Query Match      8.7%; Score 136; DB 2; Length 506;
Best Local Similarity 24.4%; Pred. No. 2.9;
Matches 94; Conservative 44; Mismatches 129; Indels 118; Gaps 21;

QY   1 AVENEKETPETPEDTS-EEEVTKANLI-FANGSTQTAETFGTPEKAATSEAYAVADTLKK 58
Db    ||| : | | | | | | | | | | | | | | | | | | | | | | | | |
Db  96 AVQSANTNSQSDLSAQEQTBRLIEDRVSGGTQ---FNQV--KVLAQDNTLTIQVGA 150
```

```
Matches 70; Conservative 39; Mismatches 126; Indels 47; Gaps 12;
Qy ✓ 11 TPTDSEEEVTIKANLIFANGSTQTAEFKGTFFKAT-SEAYAYADTLKKDNGEYTVDVAD 69
      | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 1028 TESSVAEEETK-----TTITEEVSGTSESASINSKTTMTTLLSEDTGKTSV---- 1075
Qy 70 KGYTLNKFAGKKEKTP-EPEKEEVTIKANLIYADGKTQTAEFKGTFFEEATAEAYRYADAL 128
      : : | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 1076 -SEEITTEMVTEETSETSPTEGTSKTTMTSTVSEETESS---SVTEETTTETTIVENAT 1131
Qy 129 KKONGEYTVADKGYTLNKFAGKKEKTPPEPKKEEVTIKANLIYADGKTQTAEFKGTFFEE 188
      : | | : | | : | | | | | | | | | | | | | | | | | | | |
Db 1132 DISTEVT--ASDK-TTMTTMSSESEKTEEATTEITV-----TKEVTESSST 1176
Qy 189 ATAEAYRYADLLAKENGKYTVDVADKGYTLNKFAGKKEKTPPE-----PKEEVTIKANLI 243
      || : : | : | | | | | | | | | | | | | | | | | | | | | |
Db 1177 ETATSDKTIITLSEETGK--TSAEESTTEKVTETTIVTMTPEETGKTIITSEEITIKTV- 1233
Qy 244 YADGKTQTAEFKGTFAEATAEAYRYADLLAKENGKYTDADLED 285
      | : | | | | | | | | | | | | | | | | | | | | | | | |
Db 1234 -----TEEPTDVGS-SEAITSDKTTVSTASEETGKYSVSEEE 1269
```

Search completed: October 29, 2002, 09:30:54
Job time : 29.636 secs

GenCore version 5.1.3
Copyright (c) 1993 - 2002 Compugen Ltd.

OM protein - protein search, using sw model

Run on: October 29, 2002, 09:25:26 ; Search time 7.42896 Seconds
(without alignments)
1589.653 Million cell updates/sec

Title: US-08-325-278b-1
Perfect score: 1565
Sequence: 1 AVENKEETPETDSEEV.....GGYTINIRFAGKKVDEKPEE 305

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 105224 seqs, 38719550 residues
Total number of hits satisfying chosen parameters: 105224

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : SwissProt_40:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	151.5	9.7	183	1 CNA_STAAU	Q53654 staphylococ
2	141	9.0	555	1 ED88_DAUCA	P20075 daucus caro
3	130.5	8.3	1337	1 DEXT_STRDO	P39653 streptococc
4	129	8.2	505	1 FLJB_SALTY	P52616 salmonella
5	127.5	8.1	1940	1 MYH3_CHICK	P02565 gallus gall
6	127	8.1	2459	1 MAPB_RAT	P15205 rattus norv
7	125	8.0	2468	1 MAPB_HUMAN	P46821 homo sapien
8	123.5	7.9	2464	1 MAPB_MOUSE	P14873 mus musculu
9	122	7.8	332	1 SP2B_BACSU	P37575 bacillus su
10	122	7.8	593	1 SPG2_STRSP	P19909 streptococc
11	122	7.8	1196	1 ICEV_PSEEX	O33479 pseudomonas
12	122	7.8	1200	1 ICEN_PSESY	P06620 pseudomonas
13	121.5	7.8	1500	1 SSF5_STRGN	P16952 streptococc
14	121.5	7.8	1637	1 MRSP_STAAU	P08544 staphylococ
15	121	7.7	627	1 ABPX_YEAST	Q08641 saccharomyc
16	119.5	7.6	848	1 NFM_MOUSE	P08553 mus musculu
17	117.5	7.5	650	1 NUCI_XENLA	P20397 xenopus lae
18	117.5	7.5	845	1 NFM_RAT	P12397 rattus norv
19	117.5	7.5	1433	1 SUBF_BACSU	P16397 bacillus su
20	117	7.4	1148	1 ICEK_PSEEX	O30611 pseudomonas
21	116.5	7.4	694	1 NUCI_CHICK	P15771 gallus gall
22	116.5	7.4	1935	1 MYSS_CYPCA	Q09339 cyprinus ca
23	116	7.4	915	1 NFM_HUMAN	P07197 homo sapien
24	115.5	7.4	793	1 YF06_MYCPN	P75280 mycoplasma
25	115	7.3	272	1 OSAT_BORBU	Q04968 borrelia bu
26	114.5	7.3	500	1 FLJB_SALAE	P52615 salmonella
27	114.5	7.3	1394	1 HAP_HAEIN	P45387 haemophilus
28	114.5	7.3	1409	1 HAP1_HAEIN	P44596 haemophilus
29	114	7.3	2116	1 MY52_DICDI	P08799 dictyosteli
30	112.5	7.2	300	1 SIAP_PIG	P31936 sus scrofa
31	112.5	7.2	1004	1 SLOP_BACBR	P09333 bacillus br
32	112.5	7.2	1104	1 COLA_CLOPE	P34153 clostridium
33	111.5	7.1	424	1 IRK8_MOUSE	P97794 mus musculu

RESULT 1				
ID	CNA_STAAU	STANDARD;	PRT;	1183 AA.
AC	Q53654;			
DT	15-DEC-1998	(Rel. 37, Created)		
DT	15-DEC-1998	(Rel. 37, Last sequence update)		
DT	16-OCT-2001	(Rel. 40, Last annotation update)		
DE	Collagen adhesin precursor.			
GN	CNA.			
OS	Staphylococcus aureus.			
OC	Bacteria; Firmicutes; Bacillus/Clostridium group;			
OC	Bacillus/Staphylococcus group; Staphylococcus.			
OX	NCBI_TaxID=1280;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RC	STRAIN=FDA 574;			
RC	MEDLINE=92165839; PubMed=1311320;			
RA	Patti J.M., Jonsson H., Guss B., Switalski L.M., Wiberg K.,			
RA	Lindberg M., Hoeoek M.;			
RL	J. Biol. Chem. 269:11672-11672(1994).			
[3]				
RN	[2]			
RP	ERRATUM.			
RA	Patti J.M., Jonsson H., Guss B., Switalski L.M., Wiberg K.,			
RA	Lindberg M., Hoeoek M.;			
RL	J. Biol. Chem. 269:11672-11672(1994).			
[3]				
RN	[2]			
RP	COLLAGEN-BINDING DOMAIN.			
RC	STRAIN=FDA 574;			
RC	MEDLINE=94032261; PubMed=8218209;			
RA	Patti J.M., Bolles J.O., Hoeoek M.;			
RA	"Identification and biochemical characterization of the ligand			
RT	binding domain of the collagen adhesin from Staphylococcus aureus."			
RL	Biochemistry 32:11428-11435(1993).			
[4]				
RN	[4]			
RP	X-RAY CRYSTALLOGRAPHY (2.0 ANGSTROMS) OF 169-318.			
RA	MEDLINE=97475225; PubMed=9334749;			
RA	Symersky J., Patti J.M., Carson M., House-Pompeo K., Teale M.,			
RA	Moore D., Jin L., Schneider A., DeLucas L.J., Hoeoek M.,			
RA	Narayana S.V.L.;			
RT	"Structure of the collagen-binding domain from a Staphylococcus			
RT	aureus adhesin."			
RL	Nat. Struct. Biol. 4:833-838(1997).			
CC	-!- FUNCTION: MEDIATES ATTACHMENT OF STAPHYLOCOCCAL CELLS TO			
CC	COLLAGEN-CONTAINING SUBSTRATA.			
CC	-!- SUBCELLULAR LOCATION: Type I membrane protein. Cell wall.			
CC	-!- SIMILARITY: TO OTHER STREPTOCOCCAL AND STAPHYLOCOCCAL PROTEINS			
CC	IN THE REGION OF THE MEMBRANE ANCHOR.			
CC	-----			
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration			
CC	between the Swiss Institute of Bioinformatics and the EMBL outstation -			
CC	the European Bioinformatics Institute. There are no restrictions on its			
CC	use by non-profit institutions as long as its content is in no way			
CC	modified and this statement is not removed. Usage by and for commercial			
CC	entities requires a license agreement (See http://www.isb-sib.ch/announce/			

P43597 saccharomyc
Q54114 streptococc
P48467 neurospora
Q53957 streptococc
Q62234 mus musculu
P12882 homo sapien
P05423 homo sapien
P16620 drosophila
P06176 salmonella
P13466 dictyostell
Q28641 oryctolagus
P21333 homo sapien

ALIGNMENTS

```
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; M81736; AAA20874.1; -.
DR PDB; 1AMX; 24-JUN-98.
DR InterPro; IPR001899; Gram_pos_anchor.
DR PROSITE; PS00343; GRAM_POS_ANCHORING; FALSE_NEG.
KW Signal; Repeat; Transmembrane; Cell wall; 3D-structure.
FT SIGNAL 1 29
FT CHAIN 30 1183
FT DOMAIN 30 1157
FT TRANSMEM 1158 1177
FT DOMAIN 1178 1183
FT DOMAIN 151 318
FT DOMAIN 533 1093
FT DOMAIN 1093 1157
FT DOMAIN 1151 1156
FT REPEAT 533 719
FT REPEAT 720 906
FT REPEAT 907 1093
SQ SEQUENCE 1183 AA; 133066 MW; B6A1CC072E575D76 CRC64;

Query Match
Best Local Similarity 9.7%; Score 151.5; DB 1; Length 1183;
Matches 102; Conservative 45; Mismatches 123; Indels 159; Gaps 27;

QY 2 VENKEETPETPTDSE-----EETIKANLIFANGSTQTAFFKGTPEKATS 47
Db 713 VTNK-YTPETTSISGEKVDKQDGKRPKVSV--NLLADGKVKTLTLD-----VTS 762
QY 48 EA---YAYADTLKKNG-----EYTVVADKGYTLNIFAGKEKTPPEPKEE 91
Db 763 ETNNKYEFKDLPKYDGGKIEYTVTEDHVKDYTTDI--NGTTITNKY-----TPGETSAT 815
QY 92 VT-----IKANLIYADGKTQ-----TAEFKGTFFETAFAEAYR 123
Db 816 VTKNDDNNQDKRPTKVEL-YODGKATGKTALNNSNNWTHWTGLDEKAKGQVK 874
QY 124 Y-ADALKKNGEYTV-DVDADGK-----YT-----LNIFAGKEKTPPEPKEEVTKAN 169
Db 875 YTVVELTKVGYTHVNDNDMGNLIVTNKYTPETTSISGEKVDKQDGKRPKVSYN 934
QY 170 LIYADGKTQTAEPKGFEBATAPA---YRYADLLAKENGK-----YTVVDVAD 213
Db 935 LLANGKVKTL-----DVTSETNNKYEFKDLPKYDEGKIEYTVTEDHVKDYTTDI-- 985
QY 214 KYTLNIFAGKEKTPPEPKEEVT-----IKANLIYADGKTQ----- 250
Db 986 NGTTITNKY-----TPGETSATVTKNWDNNQDKRPTKVEL-YODGKATGKTALIN 1039
QY 251 -----TAEFKGTFAEATAEAYRY-ADLLAKENGKTYADLED---GGYTINIRFAGKKVD- 300
Db 1040 ESNWTHWTGLDEKAKGQVKYTVDELTKVNG-YTHVNDNDMGNLIVTNKYTPKPKNK 1098
QY 301 -----EKPEE 305
Db 1099 PIYPERPKD 1107

RESULT 2
EDC8_DAUCA
ID EDC8_DAUCA STANDARD; PRT; 555 AA.
AC P20075;
DT 01-FEB-1991 (Rel. 17, Created)
DT 01-FEB-1991 (Rel. 17, Last sequence update)
DT 30-MAY-2000 (Rel. 39, Last annotation update)
DE Embryonic protein DC-8 (Clone 8/10).
GN DC8.
OS Daucus carota (Carrot).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
OC Asteridae; euasterids II; Apiales; Apiaceae; Daucus.
OX NCBI_TaxID=4039;
```

Db	439	MEELKQEEG	448	
		RESULT 3		
		DEXT_STRDO		
		ID	DEXT_STRDO	STANDARD; PRT; 1337 AA.
		AC	P39653;	
		DT	01-FEB-1995	(Rel. 31, Created)
		DT	01-FEB-1995	(Rel. 31, Last sequence update)
		DT	15-JUL-1998	(Rel. 36, Last annotation update)
		DE	Dextranase precursor	(EC 3.2.1.11) (Alpha-1,6-glucan-6-
		DE	glucanohydrolase).	
		GN	DEX.	
		OS	Streptococcus downei	(Streptococcus sobrinus).
		OG	Plasmid pYA902.	
		OC	Bacteria; Firmicutes; Bacillus/Clostridium group; Streptococcaceae;	
		OC	Streptococcus.	
		OX	NCBI_TaxID=1317;	
		RP	[1]	
		RP	SEQUENCE FROM N.A., AND SEQUENCE OF 31-36.	
		RC	STRAIN=6715 / UAB66;	
		RC	MEDLINE=94292401; PubMed=80211165;	
		RA	Wanda S.-Y., Curtiss R. III;	
		RT	"Purification and characterization of Streptococcus sobrinus	
		RT	dextranase produced in recombinant Escherichia coli and sequence	
		RT	analysis of the dextranase gene.";	
		RL	J. Bacteriol. 176:3839-3850(1994).	
		CC	-1- FUNCTION: MAY PLAY A ROLE IN SUCROSE-INDEPENDENT ADHERENCE TO THE	
		CC	PELICLE-COATED TOOTH SURFACE. THE ACTIVITY OF THIS ENZYME IS	
		CC	OPTIMAL AT PH 5.3 AND AT 39 DEGREES CELSIUS.	
		CC	-1- CATALYTIC ACTIVITY: Endohydrolysis of 1,6-alpha-D-glucosidic	
		CC	linkages in dextran.	
		CC	-1- SUBUNIT: HOMODIMER.	
		CC	-1- SUBCELLULAR LOCATION: Type I membrane protein. Cell wall.	
		CC	-1- SIMILARITY: TO OTHER STREPTOCOCCAL AND STAPHYLOCOCCAL PROTEINS	
		CC	IN THE REGION OF THE MEMBRANE ANCHOR.	
		CC	-1- SIMILARITY: BELONGS TO FAMILY 66 OF GLYCOSYL HYDROLASES.	
		CC	-1- CAUTION: IT IS UNCERTAIN WHETHER MET-1 OR MET-5 IS THE INITIATOR.	
		CC	-----	
		CC	This SWISS-PROT entry is copyright. It is produced through a collaboration	
		CC	between the Swiss Institute of Bioinformatics and the EMBL outstation -	
		CC	the European Bioinformatics Institute. There are no restrictions on its	
		CC	use by non-profit institutions as long as its content is in no way	
		CC	modified and this statement is not removed. Usage by and for commercial	
		CC	entities requires a license agreement (See http://www.isb-sib.ch/announce/	
		CC	or send an email to license@isb-sib.ch).	
		CC	-----	
		DR	EMBL: M96978; AAA21772.1;	
		DR	InterPro: IPR001899; Gram_pos_anchor.	
		DR	Pfam: PF00746; Gram_pos_anchor; 1.	
		DR	PROSITE: PS00343; GRAM_POS_ANCHORING; 1.	
		KW	Hydrolase; Glycosidase; Signal; Transmembrane; Repeat; Plasmid.	
		FT	SIGNAL	1 30
		FT	CHAIN	31 1337
		FT	DOMAIN	31 1337
		FT	TRANSMEM	1314 1332
		FT	DOMAIN	1333 1337
		FT	DOMAIN	1305 1310
		FT	PROTEINS.	
		FT	SEQUENCE	1337 AA; 143298 MW; B494275A77A2E3D0 CRC64;
		FT	Query Match	8.3%; Score 130.5; DB 1; Length 1337;
		FT	Best Local Similarity	22.9%; Pred. No. 1.2;
		FT	Matches	80; Conservative 51; Mismatches 133; Indels 85; Gaps 19;
		QY,	6	ETPPTPET-DSEEVTVIKANLI--FANGSTQ-----TAEFGKTFEKAISEYAYADTLKK 58
		Db	100	EQTPT--PETSASAPATSDSVEKYAQDQTQNSGPGVIRATSAQVATRSVYSS 157
		QY	59	DNGEYTVDV-ADK-----CYTLNINFAKCEKTEPEKKEVTVIKANLIYADCKTQTAEPKG 112
		Db	158	QSGDAIVDLSDAKRSYQGEDNVLNSVDFKNNTDKE--QDVTYADVYIDNKLGT----- 210

QY 6 EETPETDSEEEVTKANLIFANGSTQAEFGTSEAYADTLK-----KON 60
 Db 1698 EEPSTQNDLSLISV-SQVEASPTSSAHTPSQIASPLQE-----DTLSDVWPDRM 1750
 QY 61 GEYTVADKGYTLNIFAGKEKTPPEKKEVTK-ANLIYADG-----KTQAEFK 111
 Db 1751 SLVASEKQVSLSE-----GKUSPKSDISPLTPRESPTSPYSPGFSOSTSGAKESTAAQ 1806
 QY 112 GT-----FEATAEAYRYAD-----ALKKNGEYTVADKGYT---LNIFAGK 153
 Db 1807 TSSSPPIDAAAAPYGRSSMLFDTMQHLLALSRDLTSSVEKDNKGKTPGDFNAYQKP 1866
 QY 154 EKTPPEEKE-----EVTKANLI--YADGKTQAEFGTSEAYADTLK 203
 Db 1867 ESTTSPDEEDYDYESHEKTIQADVGVYKETER-----TIKSPDCDSGYSYETI--- 1917
 QY 204 NGKYTVADKGYTLNIFAGKEKTPPEEKE-----EVTIKANLIYADGKTQAEFGT 258
 Db 1918 -EKTTPPEGGVSCSEI-----TEKTRTPPEGGVSYEISEKT-----TRTPEVSGYT 1964
 QY 259 AEATAEAYRYADLAKENGYTADLEDGYTI 290
 Db 1965 YEKTERSRLLDDIS--NG--YDDEGGHTL 1992
 RESULT 7
 MAPB_HUMAN STANDARD; PRT; 2468 AA.
 AC P46821;
 DT 01-NOV-1995 (Rel. 32, Created)
 DT 01-NOV-1995 (Rel. 32, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Microtubule-associated protein 1B (MAP 1B) [Contains: MAP1 light chain
 Lc1].
 GN MAP1B.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Fetal brain;
 RX MEDLINE=95104835; PubMed=7806212;
 RA Lien L.L., Feener C., Fischbach N., Kunkel L.M.;
 RT "Cloning of human microtubule-associated protein 1B and the
 identification of a related gene on chromosome 15.";
 RL Genomics 22:273-280(1994).
 CC -!- FUNCTION: THE FUNCTION OF BRAIN MAPS IS ESSENTIALLY UNKNOWN.
 CC PHOSPHORYLATED MAP1B MAY PLAY A ROLE IN THE CYTOSKELETAL CHANGES
 CC THAT ACCOMPANY NEURITE EXTENSION. POSSIBLY MAP1B BINDS TO AT LEAST
 CC TWO TUBULIN SUBUNITS IN THE POLYMER, AND THIS BRIDGING OF SUBUNITS
 CC MIGHT BE INVOLVED IN NUCLEATING MICROTUBULE POLYMERIZATION AND IN
 CC STABILIZING MICROTUBULES.
 CC -!- SUBUNIT: 3 DIFFERENT LIGHT CHAINS, Lc1, Lc2 AND Lc3, CAN ASSOCIATE
 CC WITH MAP1A AND MAP1B PROTEINS.
 CC -!- DOMAIN: Has a highly basic region with many copies of the sequence
 CC KKEE and KKEI/V, repeated but not at fixed intervals, which is
 CC responsible for the binding of MAP1B to microtubules.
 CC -!- PTM: Lc1 is coexpressed with MAP1B. It is a polypeptide generated
 CC from MAP1B by proteolytic processing. It is free to associate with
 CC both MAP1A and MAP1B. It interacts with the amino-terminal region
 CC of MAP1B (by similarity).
 CC -!- SIMILARITY: TO MAP1A.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL; L06237; AAA18904.1; -

DR MIM; 157129; -
 DR InterPro; IPR000102; MAP1B_neuraxin.
 DR Pfam; PF00414; MAP1B_neuraxin; 10.
 DR PROSITE; PS00230; MAP1B_NEURAXIN; 6.
 KW Microtubules; Repeat; Phosphorylation.
 FT CHAIN ? 2468 MAP1 LIGHT CHAIN Lc1.
 FT REPEAT 1878 1894 MAP1B 1.
 FT REPEAT 1895 1911 MAP1B 2.
 FT REPEAT 1912 1928 MAP1B 3.
 FT REPEAT 1929 1945 MAP1B 4.
 FT REPEAT 1946 1962 MAP1B 5.
 FT REPEAT 1963 1979 MAP1B 6.
 FT REPEAT 1997 2013 MAP1B 7.
 FT REPEAT 2014 2030 MAP1B 8.
 FT REPEAT 2031 2047 MAP1B 9.
 FT REPEAT 2048 2064 MAP1B 10.
 FT DOMAIN 589 790 LYS-RICH (HIGHLY BASIC, CONTAINS MANY
 KKEE AND KKEI/V REPEATS).
 FT SEQUENCE 2468 AA; 270618 MW; 540839CDBF09D461 CRC64;
 Query Match 8.0%; Score 125; DB 1; Length 2468;
 Best Local Similarity 24.7%; Pred. No. 4.8;
 Matches 71; Conservative 40; Mismatches 104; Indels 72; Gaps 12;
 QY 3 ENKEETPET-----PETDSEEEVTKANLIFANGSTQAEFGTSEAYAYA 53
 Db 560 ESKEETPEVTKVNHVEKPPKVESKEKVMVK-----KDKPVKTEKPSVTEKVPs----- 609
 QY 54 DTLLKONGEYTVADKGYTLNIFAGKEKTPPEEKE-VTKANLIYADGK 104
 Db 610 ---KEEPSPVKAEVAKQATDVKPAKAEKTVKVKPEDKKEKPKKEVAKKEDK 666
 QY 105 TQ-TAEKTFTEATAEAYADLKKNGEYTVADKGYTLNIFAGKEKTPPEEKE 163
 Db 667 TPKKKEKPKKEVKVKV--K--EIKKEEKKPKKEVK-----KETPPKVEKKE 712
 QY 164 VTIKANLIYADGKTQAEFGTSEAYADLAKENGYTVDVADKGYTLNIFKA 223
 Db 713 VK-----KEEKKEKKEKPKKEIKK---LPKDAKKSTPLSEAKKPAALK-- 756
 QY 224 GKEKTPPEEKEVTKANLIYA-----DGKTQAEFGTSEAYATAEA 265
 Db 757 -----PKVPKKEESVKKDSVYAGKPKKEKGIKVIKKEGKAAEAFAAA 798
 RESULT 8
 MAPB_MOUSE STANDARD; PRT; 2464 AA.
 AC P14873;
 DT 01-APR-1990 (Rel. 14, Created)
 DT 01-APR-1990 (Rel. 14, Last sequence update)
 DT 01-MAR-2002 (Rel. 41, Last annotation update)
 DE Microtubule-associated protein 1B (MAP 1B) (MAP1.2) (MAP1(X))
 DE [Contains: MAP1 light chain Lc1].
 GN MAP1B OR MTAP1B OR MTAP5.
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.
 OX NCBI_TaxID=10090;
 RN [1]
 RP SEQUENCE FROM N.A., AND DOMAIN.
 RC STRAIN=SWISS WEBSTER; TISSUE=Brain;
 RX MEDLINE=90094539; PubMed=2480963;
 RA Noble M., Lewis S.A., Cowan N.J.;
 RT "The microtubule binding domain of microtubule-associated protein
 RT and tau.";
 RL J. Cell Biol. 109:3367-3376(1989).
 CC -!- FUNCTION: THE FUNCTION OF BRAIN MAPS IS ESSENTIALLY UNKNOWN.
 CC PHOSPHORYLATED MAP1B MAY PLAY A ROLE IN THE CYTOSKELETAL CHANGES
 CC THAT ACCOMPANY NEURITE EXTENSION. POSSIBLY MAP1B BINDS TO AT LEAST
 CC TWO TUBULIN SUBUNITS IN THE POLYMER, AND THIS BRIDGING OF SUBUNITS
 CC MIGHT BE INVOLVED IN NUCLEATING MICROTUBULE POLYMERIZATION AND IN

CC STABILIZING MICROTUBULES.
 CC -!- SUBUNIT: 3 DIFFERENT LIGHT CHAINS, LC1, LC2 AND LC3, CAN ASSOCIATE
 CC WITH MAP1A AND MAP1B PROTEINS.
 CC -!- DOMAIN: Has a highly basic region with many copies of the sequence
 CC KKEE and KKEIV, repeated but not at fixed intervals, which is
 CC responsible for the binding of MAP1B to microtubules.
 CC -!- PTM: LC1 IS COEXPRESSED WITH MAP1B. IT IS A POLYPEPTIDE GENERATED
 CC FROM MAP1B BY PROTEOLYTIC PROCESSING. IT IS FREE TO ASSOCIATE WITH
 CC BOTH MAP1A AND MAP1B. IT INTERACTS WITH THE AMINO-TERMINAL REGION
 CC OF MAP1B.
 CC -!- SIMILARITY: TO MAP1A.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
 CC or send an email to license@isb-sib.ch).
 CC
 CC EMBL; X51396; CAA35761.1; -
 CC PIR; S07549; QRMSP1.
 CC MGD; MG1:1306778; Mtap1b.
 CC InterPro: IPR000102; MAP1B_neuraxin.
 CC Pfam: PF00414; MAP1B_neuraxin; 10.
 CC PROSITE; PS00230; MAP1B_NEURAXIN; 7.
 CC Microtubules; Repeat; Phosphorylation.
 CC CHAIN ? 2464 MAP1 LIGHT CHAIN LC1.
 CC REPEAT 1874 1890 MAP1B 1.
 CC REPEAT 1891 1907 MAP1B 2.
 CC REPEAT 1908 1924 MAP1B 3.
 CC REPEAT 1925 1941 MAP1B 4.
 CC REPEAT 1942 1958 MAP1B 5.
 CC REPEAT 1959 1975 MAP1B 6.
 CC REPEAT 1993 2009 MAP1B 7.
 CC REPEAT 2010 2026 MAP1B 8.
 CC REPEAT 2027 2043 MAP1B 9.
 CC REPEAT 2044 2060 MAP1B 10.
 CC DOMAIN 589 787
 CC LYS-RICH (HIGHLY BASIC, CONTAINS MANY
 CC KKEE AND KKEI/V REPEATS).
 CC SEQUENCE 2464 AA; 270408 MW; FBD3DD99CFDDA87 CRC64;
 CC
 CC Query Match 7.9%; Score 123.5; DB 1; Length 2464;
 CC Best Local Similarity 24.7%; Pred. No. 5.9;
 CC Matches 71; Conservative 37; Mismatches 103; Indels 77; Gaps 12;
 CC
 CC 3 ENKEETPE-----TPETDSEEVITKANLIFANGSTQTAEFGTSEAYAYA 53
 CC 560 ESKEETPEVTKTSQVEKTPKVESKELVK-----KDKPVKTESKPSVTEKEVSS----- 609
 CC
 CC 54 DTLKDNGEYTVADVADKGYT-----LNKIFAGKEKTPPEKPEEVIKANLIY 100
 CC 610 ---KEQSPKVAEAKQATESPKVTQKVVKEIKTKLEEKKEKPKREVVK----- 661
 CC 101 ADGKTQTAEFGTTFEATAEAYRADAKDNNGEYTVADVADKGYTLNIFAGKEKTPPEP 160
 CC 662 -----EDKT-----PLKDEKPRKEEV-KKEIKKEIKKEERKELKKEV 698
 CC 161 KEVITKANLIYADGKTQT-AEFGTTFEATAEAYRADLAKENGYTVDV-ADKGYTL 218
 CC 699 KKETPLK-----DAKKEVKKEEKVEKKEPKKEIKKISDKIKKSTPQSDTKKPSAL 752
 CC 219 NIKFAGKEK-TPEKPEEVIKANLIYADGKTQTAEFGTTFEATAEAPA 265
 CC 753 KPVAKKEESTKKEP-----LAAGLKUDKGVKVRKKEGKTTAAATA 795
 CC
 CC RESULT 9
 CC ID SP2B_BACSU STANDARD; PRT; 332 AA.
 CC AC P37575;
 CC DT 01-OCT-1994 (Rel. 30, Created)
 CC DT 01-OCT-1994 (Rel. 30, Last sequence update)

DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Stage II sporulation protein B.
 GN SPOIIB.
 OS Bacillus subtilis.
 OC Bacteria; Firmicutes; Bacillus/Clostridium group;
 OC Bacillus/Staphylococcus group; Bacillus.
 OX NCBI_TaxID=1423;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=168 / PY79;
 RX MEDLINE=93123172; PubMed=8419299;
 RA Margolis P.S., Driks A., Losick R.;
 RT "Sporulation gene spoIIB from Bacillus subtilis.";
 RL J. Bacteriol. 175:528-540(1993).
 CC -!- FUNCTION: INVOLVED IN ENDOSPORE DEVELOPMENT.
 CC -!- INDUCTION: UNDER SPOULATION CONTROL.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
 CC or send an email to license@isb-sib.ch).
 CC
 CC EMBL; L04519; AAB59026.1; -
 CC DR EMBL; 299118; CAB14766.1; -
 CC PIR; C40646; C40646.
 CC DR Subtilist; BG10912; spoIIB.
 CC KW Sporulation; Transmembrane; Complete proteome.
 CC FT MUTAGEN 112 132 POTENTIAL.
 CC FT MUTAGEN 122 122 G->D: LOSS OF FUNCTION.
 CC SQ SEQUENCE 332 AA; 35923 MW; FBEL2F83C7569A68 CRC64;
 CC
 CC Query Match 7.8%; Score 122; DB 1; Length 332;
 CC Best Local Similarity 23.7%; Pred. No. 0.82;
 CC Matches 79; Conservative 41; Mismatches 125; Indels 88; Gaps 15;
 CC
 CC 2 VENKEET---PETPTDSEEVITKANLIFANGSTQTAEFGTSEAYAYADTLKK 58
 CC 21 INGKEETVYEQETPTEANKSMT-----FSNWEKRRQAEQVAAQSOEHP 64
 CC
 CC 59 DNGEYTVD-VADKGYTLNIFAGKEKTPPEPK--EEVTIKANLIYADGKTQTAEFGTPE 115
 CC 65 DEDEFNWDSEEDKVF-----KEDPKVVPFPQKKTKLYAKGTGAAK----PV 108
 CC 116 EATAEAYRADALKDNNGEYTVADVADKGYTLNIFAGKEKTPPEKPEEVIKANLIYADG 175
 CC 109 KRVAATTAFAAVTGTGLGLFALNIGN-----KEASAPASLEDLSGSQTAAGDTSADK 162
 CC 176 KT-----QTAEFKGTTFEATAEAYRY-----ADLLAK---ENKGYTVDA-DKGYTLN 221
 CC 163 QTSAGKQRAQTEGTYKTVAVQAGKFSNEKGAEITLQETLQETKGYSAVSLSKDDGYTVIA 222
 CC 222 FAGKEKTPPEKPEEVIKANLIYADGK-----TQTAEFGTTFEATAEAYRADLL- 272
 CC 223 GLASEKEVSQLQGVLLDSDFEAWGCKELSLSTESDWTDSFKET-AELAAKALIDEDITK 281
 CC 273 -----AKENGYKYTA-----DLED 285
 CC 282 ASVEKIEKSLGETKASETGKKEKAILQALKELED 314
 CC
 CC RESULT 10
 CC ID SPG2_STRSP STANDARD; PRT; 593 AA.
 CC AC P19909;
 CC DT 01-FEB-1991 (Rel. 17, Created)
 CC DT 01-FEB-1991 (Rel. 17, Last sequence update)
 CC DT 15-JUL-1999 (Rel. 38, Last annotation update)
 CC DE Immunoglobulin G binding protein G precursor (IGG binding protein G).
 CC GN SPG.
 CC OS Streptococcus sp. (strain G148).

OC Bacteria; Firmicutes; Bacillus/Clostridium group; Streptococcaceae;
 OC Streptococcus.
 OX NCBI_TaxID=1306;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=GI48;
 RX MEDLINE=88029445; PubMed=3665928;
 RA Olsson A., Eliasson M., Guss B., Nilsson B., Hellman U.,
 RA Lindberg M., Uhlen M.;
 RT "Structure and evolution of the repetitive gene encoding
 RT streptococcal protein G.";
 RN Eur. J. Biochem. 168:319-324(1987).
 [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=GX7805;
 RX MEDLINE=8801586; PubMed=3658689;
 RA Filpula D., Alexander P., Fahnestock S.R.;
 RT "Nucleotide sequence of the protein G gene from Streptococcus GX7805,
 RT and comparison to previously reported sequences.";
 RN Nucleic Acids Res. 15:7210-7210(1987).
 [3]
 RP SEQUENCE OF 114-593 FROM N.A.
 RC STRAIN=GI48;
 RX MEDLINE=86300657; PubMed=3017704;
 RA Guss B., Eliasson M., Olsson A., Uhlen M., Frej A.-K., Joernvall H.,
 RA Flock J.-I., Lindberg M.;
 RT "Structure of the IgG-binding regions of streptococcal protein G";
 RN EMBO J. 5:1567-1575(1986).
 [4]
 RP STRUCTURE BY NMR OF 371-427.
 RC STRAIN=GI48;
 RX MEDLINE=98290449; PubMed=9628485;
 RA Malakauskas S.M., Mayo S.L.;
 RT "Design, structure and stability of a hyperthermophilic protein
 RT variant.";
 RN Nat. Struct. Biol. 5:470-475(1998).
 CC -!- SUBCELLULAR LOCATION: Type I membrane protein. Cell wall.
 CC -!- SIMILARITY: TO OTHER STREPTOCOCCAL AND STAPHYLOCOCCAL PROTEINS
 CC IN THE REGION OF THE MEMBRANE ANCHOR.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL; X06173; CAA29540.1; -;
 DR EMBL; Y00428; CAA68489.1; -;
 DR EMBL; X04015; CAA27638.1; -;
 DR PIR; S00128; S00128.
 DR PIR; A26314; A26314.
 DR PDB; 1GB4; 22-JUL-98.
 DR InterPro; IPR002988; GA.
 DR InterPro; IPR001899; Gram_pos_anchor.
 DR InterPro; IPR000724; IgG_bind_B.
 DR Pfam; PF01468; GA; 3.
 DR Pfam; PF00746; Gram_pos_anchor; 1.
 DR Pfam; PF01378; IgG_binding_B; 3.
 DR PRINTS; PR00015; GPOSANCHOR.
 DR PROSITE; PS00343; GRAM_POS_ANCHORING; 1.
 KW IgG-binding protein; Repeat; Transmembrane; Cell wall; Signal;
 KW 3D-structure.
 FT SIGNAL 1 33
 FT CHAIN 34 593
 FT DOMAIN 34 567
 FT TRANSMEM 568 588
 FT DOMAIN 589 593
 FT CYTOPLASMIC (POTENTIAL).
 FT ALA-RICH 111
 FT DOMAIN 104 290
 FT 3 X 37 AA REPEATS.
 FT REPEAT 104 140
 FT REPEAT 179 215
 FT 1-1.
 FT 1-2.

FT REPEAT 254 290
 FT DOMAIN 303 427
 FT REPEAT 303 357
 FT REPEAT 373 427
 FT REPEAT 531 555
 FT DOMAIN 559 564
 FT PROTEINS.
 SQ SEQUENCE 593 AA; 63291 MW; 048BA760D5B2920 CRC64;
 Query Match 7.8%; Score 122; DB 1; Length 593;
 Best Local Similarity 27.9%; Pred. No. 1.5;
 Matches 72; Conservative 31; Mismatches 87; Indels 68; Gaps 18;
 QY 12 PETDSEEVTIKANLIFANGSTQTAEFKGTFFEKATSEAYAYADTK-----KDN---GE 62
 DB 299 PKTDTYK-----LILNGKT---LKG---ETTTEAVDAATAEKVKFYQYANDNGVDGE 343
 QY 63 YIVDVADKGYTLNIK---FAGKEKTPPEEPKEVTKANLIYADGKTQTAEFGKTFEEATA 119
 DB 344 WYDDATKTKFTVTEKPEVIDASELTPAVTYKLV-----NGKT---LKG---ETTT 389
 QY 120 EAYRYADALK-----KDN---GEYTDVADKGYTLNIK---FAGKEKTPPEEPKEVTK 167
 DB 390 EAVDAATAEKVKFYQYANDNGVDGEWYDDATKTKFTVTEKPEVIDASELTPAVTYKLVIN 449
 QY 168 ANLIYADGKTQTAEFKGTFFEEATAEAYRYADLAKEN---GKYTDVADKGYTLN---IK 221
 DB 450 GKTL--KGETTT---KAVDAETAETAKAFKQ---YANDNGVDGWYTDATKTKFTVTEMVTE 501
 QY 222 FAGKEKTPPEEPKEVTK 238
 DB 502 VPGDAPTEPEKPEASIP 519
 RESULT 11
 ICEV_PSESX
 ID ICEV_PSESX STANDARD; PRT; 1196 AA.
 AC 033479;
 DT 15-DEC-1998 (Rel. 37, Created)
 DT 15-DEC-1998 (Rel. 37, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Ice nucleation protein.
 GN INAV.
 OS Pseudomonas syringae.
 OC Bacteria; Proteobacteria; gamma subdivision; Pseudomonadaceae;
 OC Pseudomonas.
 OX NCBI_TaxID=317;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=INA5;
 RX MEDLINE=97462815; PubMed=9323042;
 RA Schmid D., Pridmore D., Capitani G., Battistutta R., Neeser J.-R.,
 RA Jann A.;
 RT "Molecular organisation of the ice nucleation protein Inav from
 RT Pseudomonas syringae.";
 RL FEBS Lett. 414:590-594(1997).
 CC -!- FUNCTION: ICE NUCLEATION PROTEINS ENABLE BACTERIA TO NUCLEATE
 CC CRYSTALLIZATION IN SUPERCOOLED WATER.
 CC -!- SUBUNIT: MEMBRANE ENVIRONMENT OR AGGREGATION SEEMS TO BE REQUIRED
 CC FOR ICE NUCLEATION ACTIVITY.
 CC -!- SUBCELLULAR LOCATION: Outer membrane (By similarity).
 CC -!- DOMAIN: CONTAINS MANY IMPERFECT REPEATS OF THE CONSENSUS
 CC OCTAPEPTIDE A-G-Y-G-S-T-L-T; FURTHER ON A 16-RESIDUE AND A
 CC REGIONAL 48-RESIDUE PERIODICITY IS SUPERIMPOSED.
 CC -!- SIMILARITY: BELONGS TO THE BACTERIAL ICE NUCLEATION PROTEIN
 CC FAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----


```
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: AJ001086; CAA04521.1; -.
CC HSP: P06620; 11NA.
CC InterPro: IPR000258; Ice_nucleatn.
CC Pfam: PF00818; Ice_nucleation; 61.
CC PRINTS: PR00327; ICENUCLEATN.
CC PROSITE: PS00314; ICE_NUCLEATION; 42.
CC Ice nucleation; Repeat; Outer membrane.
CC FT DOMAIN 172 1147 OCTAPEPTIDE PERIODICITY.
CC SEQUENCE 1196 AA; 117991 MW; C9E9974CB1731E68 CRC64;

Query Match 7.8%; Score 122; DB 1; Length 1196;
Best Local Similarity 25.0%; Pred. No. 3.3;
Matches 77; Conservative 33; Mismatches 110; Indels 88; Gaps 14;

QY 24 ANLIFANGSTQTAEFK-----GTEKA-----TSEAYAYADTLKDNCEYT 64
Db 712 SSLIAGYGSTQTAGFKSIMTAGYGSTQTAQERSDLVAGYGSTSTA-GYSSSLIAGYG--S 768
QY 65 VDVAADKGYTLNIFKAGKEKTPPEPKKEVT-----IKANLIYADKGTQTAEFKGTFE 115
Db 769 TOTAGYGSILTTGY-GSTQTAQENSSLTTCYSTAGYSSSLIAGYGSTQTAGYESTL- 826
QY 116 EATAEAYRYADALKKONGEYTVDVADKGYTLNIFKAGKEKTPPEPKKEVT-----I 166
Db 827 --TA-----GY-----GSTQTAQERSDLVTGYGSTSTAGY 854
QY 167 KANLIYADKGTQTAEFKGTFEATAEAYRYADLLAKENGKTYVDVADKGYTLNIFKAGKE 226
Db 855 ASLLIAGYGSTQTAGY-----ESTLTAGYGSTQTAQENSSLTTCY---GSTAGFASLL 906
QY 227 KTPPEPKKEEYTIKANLIYADKGTQTAEFKGTFAE---ATAEAYRYADLLAKENGKTYADL 283
Db 907 IAGYGSTQTAGYKSTLTLAGYGSTQTAQERSDLTLAGYGSTATAGQDSSLIAGYGSTLTSGI 966
QY 284 ED---GGY 288
Db 967 RSLTAGY 974

RESULT 12
ID ICEN_PSEY STANDARD; PRT; 1200 AA.
AC P06620;
CT 01-JAN-1988 (Rel. 06, Created)
DT 01-JAN-1988 (Rel. 06, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Ice nucleation protein.
GN INAZ.
OS Pseudomonas syringae (pv. syringae).
OC Bacteria; Proteobacteria; gamma subdivision; Pseudomonadaceae;
OC Pseudomonas
OX NCBI_TaxID=321;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=S203;
RA Green R.L., Warren G.J.;
RT "Physical and functional repetition in a bacterial ice nucleation
RL gene."; Lindow S.E., 1985.
RN [2]
RP SUBCELLULAR LOCATION.
RX MEDLINE=92297969; PubMed=2520825;
RA Lindow S.E., Lahue E., Govindarajan A.G., Panopoulos N.J., Gies D.;
RT "Localization of ice nucleation activity and the icec gene product in
RL Pseudomonas syringae and Escherichia coli.";
RN [3]
RP Mol. Plant Microbe Interact. 2:262-272(1989).
RP 3D-STRUCTURE MODELING OF 490-535.
RX MEDLINE=93360260; PubMed=8355267;
RA Kajava A.V., Lindow S.E.;
RT "A model of the three-dimensional structure of ice nucleation
Bacteria; Firmicutes; Bacillus/Clostridium group; Streptococcaceae;
```

Streptococcus.
NCBI_TaxID=1302;
[1]
RN SEQUENCE FROM N.A.
RC STRAIN=M5;
RX MEDLINE=90236997; PubMed=2185241;
RA Demuth D.R., Golub E.E., Malamud D.;
RT "Streptococcal-host interactions. Structural and functional analysis
of a Streptococcus sanguis receptor for a human salivary
glycoprotein.";
RL J. Biol. Chem. 265:7120-7126(1990).
[2]
RN REVISIONS, SEQUENCE FROM N.A.
RC STRAIN=M5;
RX MEDLINE=96332377; PubMed=8733238;
RA Demuth D.R., Duan Y., Brooks W., Holmes A.R., McNab R.,
RJ Jenkinson H.F.;
RT "Tandem genes encode cell-surface polypeptides SspA and SspB which
mediate adhesion of the oral bacterium Streptococcus gordonii to
human and bacterial receptors.";
RL Mol. Microbiol. 20:403-413(1996).
CC -1- FUNCTION: MAY BIND SALIC ACID. RESIDUES OF SALIVARY AGGLUTININ
(SAG) IN A CALCIUM-DEPENDENT REACTION. THE INTERACTION OF SAG WITH
ITS RECEPTOR IN VARIOUS ORAL STREPTOCOCCI MODULATE BACTERIAL
COLONIZATION OF ORAL TISSUE AND IS ASSOCIATED WITH REDUCED LEVELS
OF DENTAL CARIES.
CC -1- SUBCELLULAR LOCATION: Type I membrane protein. Cell wall.
CC -1- DOMAIN: THE PR2 REGION, BY SIMILARITY WITH THE PROLINE RICH
DOMAINS OF THE S.PYOGENES M6 PROTEIN AND STAPHYLOCOCCAL PROTEIN A,
MAY TRAVERSE THE CELL WALL PEPTIDOGLYCAN AND IS FOLLOWED BY
HYDROPHOBIC AA WHICH MAY FUNCTION TO ANCHOR THE PROTEIN TO THE
MEMBRANE.
CC -1- SIMILARITY: TO OTHER STREPTOCOCCAL AND STAPHYLOCOCCAL PROTEINS
IN THE REGION OF THE MEMBRANE ANCHOR.
CC -1- SIMILARITY: BELONGS TO THE SPAP/SSP5/SPAA FAMILY.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
or send an email to license@isb-sib.ch).

EMBL; U40026; AAC44100.1; --
PIL; A35186; A35186.
InterPro: IPR001899; Gram_pos_anchor.
DR Pfam; PF00746; Gram_pos_anchor; 1.
DR PROSITE; PS00343; GRAM_POS_ANCHORING; 1.
KW Signal; Repeat; Calcium-binding; Transmembrane.
FT SIGNAL 1 38
FT CHAIN 39 1500 AGGLUTININ RECEPTOR.
FT DOMAIN 39 1474 EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 1475 1495 POTENTIAL.
FT DOMAIN 1496 1500 CYTOPLASMIC (POTENTIAL).
FT SIMILAR 164 470 TO M PROTEIN OF S.PYOGENES.
FT DOMAIN 164 470 4 X APPROXIMATE TANDEM REPEATS, HR1.
FT REPEAT 164 241 1.
FT REPEAT 242 323 2.
FT REPEAT 324 405 3.
FT REPEAT 406 470 4.
FT DOMAIN 771 887 PRO-RICH (PR2).
FT DOMAIN 1414 1436 CONSERVED IN GRAM-POSITIVE COCCI SURFACE
FT DOMAIN 1467 1472 PROTEINS.
FT CA_BIND 220 235 POTENTIAL.
FT CA_BIND 301 316 POTENTIAL.
FT CA_BIND 931 950 POTENTIAL.
FT CA_BIND 1300 1315 POTENTIAL.
SQ SEQUENCE 1500 AA; DCF190E7D4D889F CRC64;
Query Match 7.8%; Score 121.5; DB 1; Length 1500;
Best Local Similarity 25.3%; Pred. No. 4.5;

Matches 76; Conservative 35; Mismatches 134; Indels 55; Gaps 13;
QY 3 ENKEETPETDSEEEVTTKANLIFANGSTQTAEEFGTPEKATSEAYAYADTLKKDNGE 62
DB 114 ETKDKGTATTATDQAQ-----KQDEIKSDYAKQAEIEITTTTEAYKKEVAHQAEATDKINAE 169
QY 63 YTVADVADKGYTLNKFAGKEKTPPEEKPEEVTIKANLIYADCKTQTAFKGTFFEEATAEAY 122
DB 170 N--KAADKYQKDL-----KSHQEEVEKI-----NTANATAKAEYEAKLAQYQ 210
QY 123 RYDALAKDNGEYTVADVADKGYTLNKFAGKEKTPPEKE--EVTIKANLIYADG----- 175
DB 211 KDLATVKKANEDSQDYQNKLSAYQTELARVQKANAKEAYEKAVKENTAKNAKVEN 270
QY 176 --KTQTAEEFGTFFEEATAEAYRVADLLA--KENGKTYTVADVADKGYTLNKFAGKEKTP 230
DB 271 EAIKQNETAKATY--EAMKQYE--ADLAAIKKANEDNDADYQAKLAAYQTELARVQKANA 328
QY 231 EPKE--EVTIKANLIYADGKTQTAEEFGTPEAEATAEAYRVADLIKEN-----GKYTADL 283
DB 329 EAKEAYDKAVKEN-----TA--KNTAIQAEANEAIKORNETAKATYDAAYKKYEADL 377
RESULT 14
MRSP_STAAU STANDARD; PRT; 1637 AA.
AC PR0544; Q92F62;
DT 01-FEB-1996 (Rel. 33, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 01-MAR-2002 (Rel. 41, Last annotation update)
DE Methicillin-resistant surface protein precursor.
GN PLS.
OS Staphylococcus aureus.
OC Bacteria; Firmicutes; Bacillus/Clostridium group.
OC Bacillus/Staphylococcus group; Staphylococcus.
OX NCBI_TaxID=1280;
RN [1]
RC SEQUENCE FROM N.A.
RP STRAIN=ISOLATE 1061;
RA Savolainen K., Kuusela P., Paulin L., Korhonen T.K.;
RT "Pis, a large repeat-rich surface protein of methicillin resistant
Staphylococcus aureus.";
RL Submitted (DEC-1998) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE OF 624-628; 676-682; 938-948; 1156-1168; 1176-1185;
RP 1199-1205 AND 1217-1224.
RC STRAIN=ISOLATE 1061;
RX MEDLINE=96270743; PubMed=8665912;
RA Hilden P., Savolainen K., Tynnelae J., Vuontola M., Kuusela P.;
RT "Purification and characterization of a plasmin-sensitive surface
protein of Staphylococcus aureus.";
RL Eur. J. Biochem. 236:904-910(1996).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
or send an email to license@isb-sib.ch).

EMBL; AF115379; AAD09131.1; --
DR InterPro: IPR001899; Gram_pos_anchor.
DR Pfam; PF00746; Gram_pos_anchor; 1.
DR PROSITE; PS00343; GRAM_POS_ANCHORING; FALSE_NEG.
KW Cell wall; Transmembrane; Antibiotic resistance; Glycoprotein; Repeat;
FT SIGNAL 1 48 POTENTIAL.
FT CHAIN 49 1637 METHICILLIN-RESISTANT SURFACE PROTEIN.
FT DOMAIN 1301 1582 141 X 2 AA TANDEM REPEATS OF D-[SAG].
FT DOMAIN 1598 1603 CONSERVED IN GRAM-POSITIVE COCCI SURFACE
FT SEQUENCE 1637 AA; 174573 MW; 75BE9ADB469BD309 CRC64;

Query Match 7.8%; Score 121.5; DB 1; Length 1637;
Best Local Similarity 27.3%; Pred. No. 4.9;
Matches 79; Conservative 20; Mismatches 111; Indels 79; Gaps 12;

QY 3 ENKEETPETDSEEEVTKANLIIFANGSTQTAFFKGTPEKATSEAYAYADTLKKNNGE 62
DB 87 EQVDVTKDTTEQASTEE---KAN-----TTEQASTEEKADTTTEQATTEAPKAEG- 133
QY 63 YTVADVADKGYTLNKFAGKEKTPPEEKVEVTKANLIYADGKTQTAFFKGTFFEATAEAY 122
DB 134 -----TDKVVETEAPKAETDKAT-----TEAPKAETDKATEAP 170
QY 123 RYADALKKONGEYTVADVADKGYTLNKF---GKEKTPPEEKVEVTKANLIYADGKTQTAFFKGTFFEATAEAY 174
DB 171 K-----TEETDKATTEAPAPAEETSKAATEAPKAETSKAATEAPKAET 217
QY 175 GKTQTAFFKGTFFEE---ATAEAYRYADLLAKENGKTVTVADVADKGYTLNKFAGKEKTPPEE 231
DB 218 EKTAETEAPKTEETDKVETEAPK-----AEETSKAATEAPKAETN-----KVETEAP 267
QY 232 PKEEVTKANLIYADGKTQTAFFKGTFAEATA-----EAYRYADLLAKE 275
DB 268 PAAEETNKA-----ATEETPAVEDTNKSNNAOPSETERTQVVVDTVAKD 312

RESULT 15
ABPX_YEAST STANDARD; PRT; 627 AA.
AC Q08641; Q08644;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 01-MAR-2002 (Rel. 41, Last annotation update)
DE Actin-binding protein ABP140.
GN ABP140 OR YOR239W/YOR240W.
OS Saccharomyces cerevisiae (Baker's yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Saccharomycetaceae; Saccharomycetes.
OX NCBI_TaxID=4932;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=S288C / FYJ1679;
RX MEDLINE=97127829; PubMed=8972580;
RA Boyer J., Michaux G., Fairhead C., Gaillon L., Dujon B.;
RT "Sequence and analysis of a 26.9 kb fragment from chromosome XV of
the yeast Saccharomyces cerevisiae.";
RL Yeast 12:1575-1586(1996).
RN [2]
RP PARTIAL SEQUENCE FROM N.A., SEQUENCE OF 1-12; 38-72; 102-115; 244-286;
RP 560-574 AND 598-608, FUNCTION, AND SUBCELLULAR LOCATION.
RC STRAIN=B35457;
RX MEDLINE=98127445; PubMed=9467951;
RA Asakura T., Sasaki T., Nagano F., Sato A., Obaishi H., Nishiohara H.,
RA Imamura H., Hotta K., Tanaka K., Nakanishi H., Takai Y.;
RT "Isolation and characterization of a novel actin filament-binding
protein from Saccharomyces cerevisiae.";
RL Oncogene 16:121-130(1998).
CC -!- FUNCTION: Binds F-actin and shows weak F-actin crosslinking
activity.
CC -!- SUBCELLULAR LOCATION: Cytoplasm and cortical cytoskeleton.
CC -!- MISCELLANEOUS: A RIBOSOMAL FRAMESHIFT OCCURS BETWEEN THE CODONS
FOR LEU-276 AND GLY-277.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; 275147; CAA99460.1; ALT_SEQ.
DR EMBL; 275147; CAA99461.1; ALT_SEQ.

DR SGD; S0005765; ABP140.
DR InterPro; IPR001601; Meth-transf.
DR InterPro; IPR000051; SAM_bind.
KW Actin-binding; Cytoskeleton; Ribosomal frameshift.
FT INIT_MET 0
SQ SEQUENCE 627 AA; 71354 MW; D4E55F9485412F39 CRC64;

Query Match 7.7%; Score 121; DB 1; Length 627;
Best Local Similarity 22.5%; Pred. No. 1.9;
Matches 64; Conservative 52; Mismatches 103; Indels 66; Gaps 12;

QY 4 NKEETPETDSEEEVTKANLIIFANGSTQTAFFKGTPEKATSEAYAYADTLKK-DNGE 62
DB 13 SKEEGDATVDTNSSK-PLKS-----DETRELHQESTAVPQEVDAVEEFENEP 61
QY 63 YTVD---VADKGYTLNKFAGKEKTPPEEKVEVTKANLIYADGKT-----QTAEF 110
DB 62 ETINSRTAEKPLETNLP---KPEETNEDEEEGSMSENKIYSGENADINVNDFOEYKEM 118
QY 111 KGTFFEATAEAYRYADALKKDNGEYTVADVADKGYTLNKFAGKEKTPPEEKVEVTKANL 170
DB 119 ENTGAELASSVEESDAIQEGVAETEGIA-----TPKOKENEKNDESEE 163
QY 171 IYADGKTQTA-EFKGTFFEATAEAYRYADLLAKENGKTYVDV-----ADKGYT----L 218
DB 164 ESANNASEPAEYSQSEEDADIE-----QSNKGETENAENASQOANDGSTSTTSK 214
QY 219 NIKFAGKEKTPPEEKVEVTKANLIYADGKTQTAFFKGTFAEATA 263
DB 215 NKKKKNNKKNNKRRNGVNTNANV---DDSTKTGENDDTTGDTTTS 256

Search completed: October 29, 2002, 09:29:54
Job time : 14.429 secs

GenCore version 5.1.3
Copyright (c) 1993 - 2002 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: October 29, 2002, 09:27:36 ; Search time 12.7943 Seconds
(without alignments)
2290.646 Million cell updates/sec

Title: US-08-325-278B-1

Perfect score: 1565

Sequence: 1 AVENKEETPETDSEEV.....GGYTINIRFAGKVKDEKPEE 305

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 283138 seqs, 96089334 residues

Total number of hits satisfying chosen parameters: 283138

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

PIR_71.*
1: pir1.*
2: pir2.*
3: pir3.*
4: pir4.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1565	100.0	455	A45063	immunoglobulin-bin
2	1561	99.7	719	A42808	Ig light chain-bin
3	1326.5	78.4	932	S54396	protein L precursor
4	171.5	11.0	74	A34483	Ig light chain-bin
5	167	10.7	664	S42574	streptococcal surf
6	166	10.6	479	T47561	late embryogenesis
7	151.5	9.7	1185	A42404	collagen adhesin -
8	151	9.6	448	H84782	late embryogenesis
9	149	9.5	448	JC6171	late embryogenesis
10	141.5	9.0	1110	I51116	NF-180 - sea lamp
11	141	9.0	555	S04309	embryonic protein
12	137.5	8.8	992	E72350	hypothetical prote
13	132.5	8.5	737	H81070	lactoferrin-bindin
14	131.5	8.4	891	G41662	130K surface exclu
15	131.5	8.4	1166	T28680	fibrinogen-binding
16	130.5	8.3	701	H98120	choline binding pr
17	130.5	8.3	1141	E98824	hypothetical prote
18	130.5	8.3	1337	T30291	dextranase - Strep
19	128.5	8.2	938	AF1772	internalin-like pr
20	127.5	8.1	843	S72375	surface exclusion
21	127.5	8.1	1940	A29320	myosin heavy chain
22	127	8.1	863	S06017	neuraxin - rat
23	127	8.1	913	T52485	neurofilament prot
24	127	8.1	2364	A56577	microtubule-associ
25	126.5	8.1	1035	T16588	hypothetical prote
26	126	8.1	955	A60990	alpha-amylase (EC
27	125.5	8.0	673	C86278	F14L17.15 protein
28	125.5	8.0	1536	A45605	mature-parasite-in
29	125	8.0	463	S27757	embryonic abundant

surface-located me
hypothetical prote
cyst germination s
ladder protein - C
microtubule-associ
hypothetical prote
EF protein - Strep
myosin heavy chain
surface exclusion
myosin heavy chain
probable peptidogl
endospore developm
hypothetical prote
protein G precursor
ice nucleation pro
probable cell surf

30 124.5 8.0 1302 1 JC6009
31 124 7.9 495 2 D89808
32 124 7.9 1489 2 T31108
33 123.5 7.9 1198 2 T42223
34 123.5 7.9 2464 1 QRMSPI
35 123 7.9 569 2 T39577
36 123 7.9 1822 2 S33441
37 122.5 7.8 698 2 S52696
38 122.5 7.8 890 2 S22452
39 122.5 7.8 955 2 S24348
40 122.5 7.8 1993 2 AF1450
41 122 7.8 332 2 C40646
42 122 7.8 384 2 G86287
43 122 7.8 593 2 S00128
44 122 7.8 1200 1 SNPSO
45 121.5 7.8 657 2 AD1525

ALIGNMENTS

RESULT 1

A45063

immunoglobulin-binding protein LG - Peptostreptococcus magnus

C:Species: Peptostreptococcus magnus

C:Date: 30-Apr-1993 #sequence_revision 18-Nov-1994 #text_change 15-Oct-1999

C:Accession: A45063

R:Kihlberg, B.M.; Sjobring, U.; Kastern, W.; Bjorck, L.

J. Biol. Chem. 267, 25583-25588, 1992

A:Title: Protein LG: a hybrid molecule with unique immunoglobulin binding properties.

A:Reference number: A45063; MUID:93094283

A:Accession: A45063

A:Status: preliminary

A:Molecule type: mRNA; protein

A:Residues: 1-455 <Kih>

A:Cross-references: GB:S50809; NID:g261705; PIDN:AAA03280.1; PID:g261706

A:Note: sequence extracted from NCBI backbone (NCBIN:120302, NCBIP:120303)

Query Match 100.0%; Score 1565; DB 2; Length 455;

Best Local Similarity 100.0%; Pred. No. 2e-85;

Matches 305; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 AVENKEETPETDSEEVTIKANLIFANGSTQTAEFKGTFFEKATSEAYAYADTLKKDN 60
|||||
Db 22 AVENKEETPETDSEEVTIKANLIFANGSTQTAEFKGTFFEKATSEAYAYADTLKKDN 81
Qy 61 GEYTVADVADKGYTLNIFAGKEKTPPEPKEVTTKANLIYADGKTQTAEFKGTPEEATAE 120
|||||
Db 82 GEYTVADVADKGYTLNIFAGKEKTPPEPKEVTTKANLIYADGKTQTAEFKGTPEEATAE 141
Qy 121 AYRYADALKDNGEYTVADVADKGYTLNIFAGKEKTPPEPKEVTTKANLIYADGKTQTA 180
|||||
Db 142 AYRYADALKDNGEYTVADVADKGYTLNIFAGKEKTPPEPKEVTTKANLIYADGKTQTA 201
Qy 181 EFKGTPEEATAEAYRYADLLAKENGYTVADVADKGYTLNIFAGKEKTPPEPKEVTTKAN 240
|||||
Db 202 EFKGTPEEATAEAYRYADLLAKENGYTVADVADKGYTLNIFAGKEKTPPEPKEVTTKAN 261
Qy 241 NLIYADGKTQTAEPKGTFAEATAEAYRYADLLAKENGYTADLEDGGYTTINIRFAGKVD 300
|||||
Db 262 NLIYADGKTQTAEPKGTFAEATAEAYRYADLLAKENGYTADLEDGGYTTINIRFAGKVD 321
Qy 301 EKPEE 305
Db 322 EKPEE 326

RESULT 2

A42808

Ig light chain-binding protein precursor - Peptostreptococcus magnus

N:Alternate names: protein L

C:Species: Peptostreptococcus magnus

C:Date: 19-May-1994 #sequence_revision 19-May-1994 #text_change 07-May-1999

C;Accession: A42808; A41493
R;Kastern, W.; Sjoerding, U.; Bjoerck, L.
J. Biol. Chem. 267, 12820-12825, 1992
A;Title: Structure of peptostreptococcal protein L and identification of a repeated immun
A;Reference number: A42808; MUID:92316971
A;Accession: A42808
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-719 <RAS>
A;Cross-references: GB:M86697
R;Kastern, W.; Holst, E.; Nielsen, E.; Sjoerding, U.; Bjoerck, L.
Infect. Immun. 58, 1217-1222, 1990
A;Title: Protein L, a bacterial immunoglobulin-binding protein and possible virulence de
A;Reference number: A41493; MUID:90215984
A;Accession: A41493
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 202-275 <KA2>
C;Keywords: immunoglobulin

Query Match 99.7%; Score 1561; DB 2; Length 719;
Best Local Similarity 100.0%; Pred. No. 5.8e-85;
Matches 304; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 VENKEETPETDSEEVITKANLIFANGSTQTAEPKGFERKATSEAYAYADTLKKDNG 61
Db 95 VENKEETPETDSEEVITKANLIFANGSTQTAEPKGFERKATSEAYAYADTLKKDNG 154

QY 62 EYTVADVADKGYTLNIFAGKEKTPPEPKEEVTIKANLIYADGKTQTAEPKGFEEATAEA 121
Db 155 EYTVADVADKGYTLNIFAGKEKTPPEPKEEVTIKANLIYADGKTQTAEPKGFEEATAEA 214

QY 122 YRYADALKKONGEYTVADVADKGYTLNIFAGKEKTPPEPKEEVTIKANLIYADGKTQTAEP 181
Db 215 YRYADALKKONGEYTVADVADKGYTLNIFAGKEKTPPEPKEEVTIKANLIYADGKTQTAEP 274

QY 182 FKGFEEATAEAAYRYADLLAKENGKYTVADVADKGYTLNIFAGKEKTPPEPKEEVTIKAN 241
Db 275 FKGFEEATAEAAYRYADLLAKENGKYTVADVADKGYTLNIFAGKEKTPPEPKEEVTIKAN 334

QY 242 LIYADGKTQTAEPKGFPAEATAEAYRYADLLAKENGKYTVADVADGKYTLNIRFAGKKVDE 301
Db 335 LIYADGKTQTAEPKGFPAEATAEAYRYADLLAKENGKYTVADVADGKYTLNIRFAGKKVDE 394

QY 302 KPEE 305
Db 395 KPEE 398

RESULT 3
S54396
protein L precursor - Peptostreptococcus magnus (strain 3316)
C;Species: Peptostreptococcus magnus
A;Variety: strain 3316
C;Date: 27-Oct-1996 #sequence_revision 13-Mar-1997 #text_change 15-Oct-1999
C;Accession: S54396
R;Murphy, J.P.; Duggleby, C.J.; Atkinson, M.A.; Trowern, A.R.; Atkinson, T.; Goward, C.H
Mol. Microbiol. 12, 911-920, 1994
A;Title: The functional units of a peptostreptococcal protein L.
A;Reference number: S54396; MUID:95020613
A;Accession: S54396
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-992 <MUR>
A;Cross-references: EMBL:L04466; NID:g150673; PIDN:AAA67503.1; PID:g150674

Query Match 78.4%; Score 1226.5; DB 2; Length 992;
Best Local Similarity 79.9%; Pred. No. 4.6e-65;
Matches 246; Conservative 22; Mismatches 31; Indels 9; Gaps 4;

QY 2 VENKEETPETDSEEVITKANLIFANGSTQTAEPKGFERKATSEAYAYADTLKKDNG 61
Db 241 MERKLEKETPE--PEEVTIKANLIFADGSGTQNAEPKGFPAKAVSDAYAYADALKKNG 298

QY 62 EYTVADVADKGYTLNIFAGKEKTPPEPKEEVTIKANLIYADGKTQTAEPKGFEEATAEA 121
Db 299 EYTVADVADKGLNINIFAGKKEKPEPKEEVIKVNLIYFADGKTQTAEPKGFEEATAEA 358

QY 122 YRYADALKKONGEYTVADVADKGYTLNIFAGKE--KTPEEPKEEVTIKANLIYADGKTQT 179
Db 359 YAYADLLAKENGEXTADLEGGNTINIFAGKETPETPEPKEEVTIKVNLIYFADGKIQT 418

QY 180 ABEKGTFFEEATAEAAYRYADLLAKENGKYTVADVADKGYTLNIFAGKE--KTPEEPKEEVT 237
Db 419 ABEKGTFFEEATAKAYAYANLLAKENGEXTADLEGGNTINIFAGKETPETPEPKEEVT 478

QY 238 IKANLIYADGKTQTAEPKGFPAEATAEAYRYADLLAKENGKYTVADLEGGYTLNIRFAGK 297
Db 479 IKVNLIYFADGKTQTAEPKGFPAEATAEAYRYADLLAKVNGEXTADLEGGYTLNIRFAGK 538

QY 298 KVDEKPEE 305
Db 539 ---EQPGE 543

RESULT 4
A34483
Ig light chain-binding protein L - Peptostreptococcus magnus (fragments)
C;Species: Peptostreptococcus magnus
C;Date: 07-Sep-1990 #sequence_revision 07-Sep-1990 #text_change 16-Aug-1996
C;Accession: A34483
R;Akerstroem, B.; Bjoerck, L.
J. Biol. Chem. 264, 19740-19746, 1989
A;Title: Protein L: an immunoglobulin light chain-binding bacterial protein. Characte
A;Reference number: A34483; MUID:90062074
A;Accession: A34483
A;Molecule type: protein
A;Residues: 1-74 <AKE>
C;Keywords: immunoglobulin

Query Match 11.0%; Score 171.5; DB 2; Length 74;
Best Local Similarity 58.2%; Pred. No. 0.00038;
Matches 39; Conservative 5; Mismatches 4; Indels 19; Gaps 2;

QY 82 EKTPEPKEEVTIKANLIYADGKTQTAEPKGF-----TFEEATAEAAYYA 125
Db 9 ETTPEPKEEVTIKANLIYADGKTETAEKFGPEETPEKPEVDGQYASVEEATAAA---K 65

QY 126 DALKKDN 132
Db 66 EALKND 72

RESULT 5
S42574
streptococcal surface protein - Streptococcus dysgalactiae
C;Species: Streptococcus dysgalactiae
C;Date: 13-Jan-1995 #sequence_revision 13-Jan-1995 #text_change 19-May-2000
C;Accession: S42574
R;Jonsson, H.; Mueller, H.P.
Eur. J. Biochem. 220, 819-826, 1994
A;Title: The type-III Fc receptor from Streptococcus dysgalactiae is also an alpha(2)
A;Reference number: S42574; MUID:94192673
A;Accession: S42574
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-664 <JON>
A;Cross-references: EMBL:Z29666; NID:g470220; PIDN:CAA82764.1; PID:g470221
C;Superfamily: M5 protein

Query Match 10.7%; Score 167; DB 2; Length 664;
Best Local Similarity 27.4%; Pred. No. 0.0092;
Matches 84; Conservative 49; Mismatches 134; Indels 40; Gaps 16;

QY 7 ETPETDSEEVITKANLIFANGSTQTAEPKGFERKATSEAYAYADTLKKDNGEYTV 66
Db 241 MERKLEKETPE--PEEVTIKANLIFADGSGTQNAEPKGFPAKAVSDAYAYADALKKNG 298

Db 214 DAPELTPALTYKLIVKGN--TFSGETTTKAITATAAEKFKQ-YATANNV---DGEWSYD 268

QY 67 VADKGYTLNKAFA--GKEKTPPEKPEVTKANLIYADGKTOTAEFKGTPEEATAEAYR 123

Db 269 DAYKTFVTTEKPAVIDAPELTPALTYKLVKGNTF--SGETTT--KAVDAETAERAKFK 323

QY 124 -YADALKKONGEYTDVADKGYTLINKAFA--GKEKTPPEKPEVTKANLIYADGKTOT 179

Db 324 QYATANNVD--GEWSYDDATFTVTEKPAVIDAPELTPALTYKLVKGNTF--SGETTT 380

QY 180 AEFKGTPEEATAEAYRYADLLAKENGKYTVDDVADKGYTLINKAFA--GKEKTPPEKPEEV 236

Db 381 KAIDAATAERFKOYATANGV---DGEWSYDDATKFTVTEKPAVIDAPELTPALTYKL 437

QY 237 TIRKANLIYADGKTOTAEFKGTPEEATAEAYRYADLLAKENGKYTADLEDGKYTINIRFAG 296

Db 438 IVKGNTF--SGETTT--KAVDAETAERAKFKQ---YANENGVI-----GEWSYDDATKT 483

QY 297 KKYDEKP 303

Db 484 FTVTEKP 490

RESULT 6

T47561

N: late embryogenesis abundant protein-like - Arabidopsis thaliana

C: Species: Arabidopsis thaliana (mouse-ear cress)

C: Date: 20-Apr-2000 #sequence_revision 20-Apr-2000 #text_change 20-Apr-2000

C: Accession: T47561

R: Nyakatura, G.; Fartmann, B.; Dauner, D.; Sterr, W.; Holland, R.; Weichselgartner, M.; Mayer, K.F.X.

submitted to the Protein Sequence Database, April 2000

A: Reference number: Z24458

A: Accession: T47561

A: Status: preliminary

A: Molecule type: DNA

A: Residues: 1-479 <N>A>

A: Cross-references: EMBL:AL132969

A: Experimental source: cultivar Columbia; BAC clone F8J2

C: Genetics:

A: Map position: 3

A: Introns: 329/3; 382/2

A: Note: F8J2.210

Query Match 10.6%; Score 166; DB 2; Length 479;

Best Local Similarity 25.8%; Pred. No. 0.0072;

Matches 82; Conservative 43; Mismatches 137; Indels 56; Gaps 12;

QY 3 ENKEETPETPTDSEEVTTKANLIFANGSTQTAEFKGTPEKATSEAYAY---ADFLKK 58

Db 75 DTAESTRGADIASEKAAGR-----DTTGEVRDSTAQKTETADTADKAREAKD 125

QY 59 DNGEYTVDDVADKGYTLINIKFAGKEKTPPEKPEVTTKANLIYADGKTOTAEFKGTPEEAT 118

Db 126 KTADKTKRETAD--YAAEKAREAKDRDADTKETAETAYAEKA-REAKDKTADKLGEYKDYT 182

QY 119 AEAYRYA-DALKDNGEYTVDDVADKGYTLINIKFAGKEKTPPEKPEVTTKANLIYADGKT 177

Db 183 AEKAKEAKDTTAEKLGEY-----KDYTVDKAKEAKDKTAEKATAEYTSDKA-RETKD 235

QY 178 QTAEFKGTPEEATAE-----AYRYADLLAKENGKYTVDDVADKGYTLINIKFAGKEKTP 229

Db 236 KTAEKVGEYKDYTAEKAKETADKAREAKDKTAEKVGEYRDYTAEK--ATETKDGWSKIG 293

QY 230 EEPKEEV-TIKANLIYADGKTQ-----TAEFKGTFAEATAEAYRYADLLAKENGKYTADL 283

Db 294 ELKDSAVDTAKRANGFLSGKTEETKQAVETKDTAKEKMDGEAGEARRKKEE----- 345

QY 284 EDGKYTINIRFAGKYVDE 301

Db 346 -----MRLEGGKKLDE 355

RESULT 7

A42404

collagen adhesin - Staphylococcus aureus

C: Species: Staphylococcus aureus

C: Date: 04-Mar-1993 #sequence_revision 18-Nov-1994 #text_change 10-Nov-1995

C: Accession: A42404; S27665

R: Patti, J.M.; Jonsson, H.; Guss, B.; Switalski, L.M.; Wiberg, K.; Lindberg, M.; Hook J.; Biol. Chem. 267, 4766-4772, 1992

A: Title: Molecular characterization and expression of a gene encoding a Staphylococcus A: Reference number: A42404; MUID: 92165839

A: Contents: FDA 574

A: Accession: A42404

A: Status: preliminary

A: Molecule type: DNA

A: Residues: 1-1185 <PAT>

A: Cross-references: EMBL:M81736

A: Note: sequence extracted from NCBI backbone (NCBIP: 83982)

Query Match 9.7%; Score 151.5; DB 2; Length 1185;

Best Local Similarity 23.8%; Pred. No. 0.15;

Matches 102; Conservative 45; Mismatches 123; Indels 159; Gaps 27;

QY 2 VENKEETPETPTDSE-----EEVTIRKANLIFANGSTQTAEFKGTPEKATSEAYR 47

Db 715 VTNK-YTPETTSISGEKVVDDKDNQDKGRPEKYSV--NLLADGEKVKTLT-----VTS 764

QY 48 EA---YAYADTLKKDNG-----EYTVDDVADKGYTLINIKFAGKEKTPPEEK 91

Db 765 ETNWKYEFKDLPRYDEGKIEYTVTDHVKDYTTDI--NGTTITNKY-----TPGETSAT 817

QY 92 VT-----IRKANLIYADGKTQ-----TAEFKGTPEEATAEAYR 123

Db 818 VTNNDDNNQDKGRPEKIEKVEL-YDQKATGKTALINESNNHTWTGLDEKAKGQOVK 876

QY 124 Y-ADALKKONGEYTVDDVADK-----YT-----LNIRFAGKEKTPPEEK 169

Db 877 YTVVELTKVGYTTHVDNDMGNLIVTNKYTPETTSISGEKVVDDKDNQDKGRPEKYSV 936

QY 170 LIYADGKTQTAEFKGTPEEATAE---YRYADLLAKENGK-----YTVDVAD 213

Db 937 LLANGEKVKTL-----DVTSETNWKYEFKDLPRYDEGKIEYTVTDHVKDYTTDI-- 987

QY 214 KGYTLINIKFAGKEKTPPEEKPEVTT-----IKANLIYADGKTQ----- 250

Db 988 NGTITNKY-----TPGETSATVTNNDDNNQDKGRPEKIEKVEL-YDQKATGKTALIN 1041

QY 251 -----TAEFKGTFAEATAEAYRY-ADLLAKENGKYTADLED---GGYTIINIRFAGKKVD- 300

Db 1042 ESNWHTHTWGLDEKAKGQOVKYTVDELTKVNG-YTTHVDNDMGNLIVTNKYTPKKPNK 1100

QY 301 -----EKPEE 305

Db 1101 PIYPEKPKD 1109

RESULT 8

H84782

late embryogenesis abundant protein (AtECP63) [imported] - Arabidopsis thaliana

C: Species: Arabidopsis thaliana (mouse-ear cress)

C: Date: 02-Feb-2001 #sequence_revision 02-Feb-2001 #text_change 02-Feb-2001

C: Accession: H84782

R: Lin, X.; Kaul, S.; Rounsley, S.D.; Shea, T.P.; Benito, M.I.; Town, C.D.; Fujii, C.Y. M.; Koo, H.; Moffat, K.S.; Cronin, L.A.; Shen, M.; VanAken, S.E.; Umayam, L.; Tallon, eus, D.; Nierman, W.C.; White, O.; Eisen, J.A.; Salzberg, S.L.; Fraser, C.M.; Venter Nature 402, 761-768, 1999

A: Title: Sequence and analysis of chromosome 2 of the plant Arabidopsis thaliana.

A: Reference number: A84420; MUID: 20083487

A: Accession: H84782

A: Status: preliminary

A: Molecule type: DNA

A: Residues: 1-448 <STO>

A: Cross-references: GB:AE002093; NID:g4415909; PIDN:AAD20140.1; GSPDB:GN00139

C:Genetics:

A:Gene: At2g36640

A:Map position: 2

Query Match 9.6%; Score 151; DB 2; Length 448;

Best Local Similarity 24.5%; Pred. No. 0.051;

Matches 79; Conservative 38; Mismatches 107; Indels 98; Gaps 14;

QY 3 ENKEETPETDSEEVTTKANLIIFANGSTQTAFFKGTPEKATSEAYAYADTLKKDNGE 62

DB 79 EAAETKEGAQIASEKAVGAK-----DAIVEKAKET-----ADYTAEKVGE 119

QY 63 YTVADVADKGYTLNKKFAGKEKTPPEKPEEVIKANLIYADGKTQTAFFKGTPEATAEAY 122

DB 120 Y-----KDYTVDKAKEADTTAEKAKETANYTADKA-VEAKDKTAETKIGEYK----- 165

QY 123 RYADALKKNGEYTVADVADKGYTLNKKFAGKEKTPPEKPEEVIKANLIYADGKTQTAFF 182

DB 166 -----DYAVDKA-----VEAKDKTAETKAKETANYTADKA-KEAKDKTAEK 204

QY 183 KGTFEEATAE-----AYRYADLLAKENGKYTVADVADKGYTLNKKFAGKEKTPPE 230

DB 205 VGEYKDYTVDKAVEADYDTAEKAEIAEKDKTAETKIGY-----KDYTVKATEGKDVTVS 258

QY 231 ----EPKEEV--TIKANLIYADGKTQTAFFKGTPEATAEAYRYADLLAKENGKYTADLED 285

DB 259 KLGLKDSAVETAKRAMGFLSGKTEAKGKAVETKDT-----AKEN-----MEK 302

QY 286 GGYTI-----NIRFAGKKVDEK 302

DB 303 AGEVTRQKMEEMRLGKELKEE 324

RESULT 9

JC6171

late embryogenesis abundant protein - Arabidopsis thaliana

N:Alternate names: embryogenic cell protein 63; phosphotyrosine protein

C:Species: Arabidopsis thaliana (mouse-ear cress)

C:Date: 11-Apr-1997 #sequence_revision 09-May-1997 #text_change 21-Jul-2000

C:Accession: JC6171

R:Yang, H.; Saitou, T.; Kameda, Y.; Harada, H.; Kamada, H.

Gene 184, 83-88, 1997

A:Title: Arabidopsis thaliana ECP63 encoding a LEA protein is located in chromosome 4.

A:Reference number: JC6171; MUID:97169149

A:Accession: JC6171

A:Molecule type: mRNA

A:Residues: 1-448 <YAN>

A:CROSS-references: DDBJ:D641140; NID:gl526423; PIDN:BA11017.1; PID:gl526424

C:Comment: This protein is a phosphotyrosine protein which belongs to late embryogenesis

C:Genetics:

A:Gene: ATECP63

A:Map position: 4, south part

C:Keywords: seed

Query Match 9.5%; Score 149; DB 2; Length 448;

Best Local Similarity 24.5%; Pred. No. 0.067;

Matches 79; Conservative 38; Mismatches 107; Indels 98; Gaps 14;

QY 3 ENKEETPETDSEEVTTKANLIIFANGSTQTAFFKGTPEKATSEAYAYADTLKKDNGE 62

DB 79 EAAETKEGAQIASEKAVGAK-----DAIVEKAKET-----ADYTAEKVGE 119

QY 63 YTVADVADKGYTLNKKFAGKEKTPPEKPEEVIKANLIYADGKTQTAFFKGTPEATAEAY 122

DB 120 Y-----KDYTVDKAKEADTTAEKAKETANYTADKA-VEAKDKTAETKIGEYK----- 165

QY 123 RYADALKKNGEYTVADVADKGYTLNKKFAGKEKTPPEKPEEVIKANLIYADGKTQTAFF 182

DB 166 -----DYAVDKA-----VEAKDKTAETKAKETSNTYADKA-KEAKDKTAEK 204

QY 183 KGTFEEATAE-----AYRYADLLAKENGKYTVADVADKGYTLNKKFAGKEKTPPE 230

DB 205 VGEYKDYTVDKAVEADYDTAEKAEIAEKDKTAETKIGY-----KDYTVKATEGKDVTVS 258

QY 231 ----EPKEEV--TIKANLIYADGKTQTAFFKGTPEATAEAYRYADLLAKENGKYTADLED 285

DB 259 KLGLKDSAVETAKRAMGFLSGKTEAKGKAVETKDT-----AKEN-----MEK 302

QY 286 GGYTI-----NIRFAGKKVDEK 302

DB 303 AGEVTRQKMEEMRLGKELKEE 324

RESULT 10

IS1116

NF-180 - sea lamprey

C:Species: Petromyzon marinus (sea lamprey)

C:Date: 13-Sep-1996 #sequence_revision 13-Sep-1996 #text_change 21-Jul-2000

C:Accession: IS1116

R:Jacobs, A.J.; Kamholz, J.; Selzer, M.E.

Brain Res. Mol. Brain Res. 29, 43-52, 1995

A:Title: The single lamprey neurofilament subunit (NF-180) lacks multiphosphorylation

A:Reference number: IS1116; MUID:95287814

A:Accession: IS1116

A>Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: mRNA

A:Residues: 1-1110 <JAC>

A:CROSS-references: EMBL:U19361; NID:g632548; PIDN:AAA80106.1; PID:g632549

C:Superfamily: neurofilament triplet H protein

Query Match 9.0%; Score 141.5; DB 2; Length 1110;

Best Local Similarity 25.4%; Pred. No. 0.54;

Matches 87; Conservative 38; Mismatches 138; Indels 79; Gaps 13;

QY 1 AVENKEETPETDSEEVTTKANLIIFANGSTQTAFFKGTPE-KATSEAYAYADTLKKD 59

DB 482 SAEKDEEEEEEKEEEE-----AAEEEEEDGRKEGEAAEEEEEEVEKEE 532

QY 60 NGEYTVADVADKGYT-----LNKFAKKEKTPPEKPEEVIKANLIYADGKT 105

DB 533 AEEAEVVEEAEEAEETEAEEAEAEAEAEAEAEAEAEAEAEAEAEAEAEAEAE 584

QY 106 QTAFFKGTPEATAE-----YRYADLLKNGEYTVADVADKGYTLNKKFAGKEKT 156

DB 585 EAAEAKAEVEEEAEAEAEAEAEAEAEAEAEAEAEAEAEAEAEAEAEAE 638

QY 157 PEE--PKEEVIKANLIYADGKTQTAFFKGTPEATAEAYRYADLLAKENGKYTVADVADK 214

DB 639 EEEAEAEAEVTSK-----KAKTQAE-----VEEEAEAEAEAEAEAEAEAE 684

QY 215 GYTLNKKFAGKEKTPPEK-----EVTIKANLIYADGKTQTAFFKGTFAEATAE 264

DB 685 -----VEAESEKEEEEDSKADAEDEAEPEEEVTKSDAEAEAEAEAEAEAE 739

QY 265 AYRYA-DLLAKENGKYTADLEDGGYTTINIRFAGKKVDEKPEE 305

DB 740 AAEEAKDEAE 781

RESULT 11

S04909

embryonic protein DC8 (clone 8/10) - carrot

C:Species: Daucus carota (carrot)

C:Date: 28-Feb-1990 #sequence_revision 28-Feb-1990 #text_change 21-Jul-2000

C:Accession: S04909

R:Franz, G.; Hatzopoulos, P.; Jones, T.J.; Krauss, M.; Sung, Z.R.

Mol. Gen. Genet. 218, 143-151, 1989

A:Title: Molecular and genetic analysis of an embryonic gene, DC 8, from Daucus carot

A:Reference number: S04909; MUID:89384429

A>Status: not compared

A:Molecule type: DNA

A:Residues: 1-555 <FRA>

A:CROSS-references: GB:X16131; NID:gl18333; PIDN:CAA34258.2; PID:g4902464

Query Match

9.0%; Score 141; DB 2; Length 555;


```

QY   247  GKQTAEFKGTFAEATAEARYADLL-----AKENGKKTADLED 285
      |         |         |         |         |         |
Db   477  GVEFYVE-----RYTDRFVYMGDNVSRNGKMSFEKEE 509

RESULT 13
R81070
lactoferrin-binding protein B NMB1541 [imported] - Neisseria meningitidis (strain MC5)
C:Species: Neisseria meningitidis
C:Date: 31-Mar-2000 #sequence_revision 31-Mar-2000 #text_change 19-Jan-2001
C:Accession: H81070
R:Rettelin, H.; Saunders, N.J.; Heidelberg, J.; Jeffries, A.C.; Nelson, K.E.; Eisen,
Hickey, E.K.; Haft, D.H.; Salzberg, S.L.; White, O.; Fleischmann, R.D.; Dougherty, B.
xi, H.; Qin, H.; Vanathevan, J.; Gill, J.; Scarlato, V.; Massignani, V.; Pizzo, M.
Science 287, 1809-1815, 2000
A:Authors: Grandal, G.; Sun, L.; Smith, H.O.; Fraser, C.M.; Moxon, E.R.; Rappuoli, R.;
A:Title: Complete genome sequence of Neisseria meningitidis serogroup B strain MC58.
A:Reference number: A81000; MUID:20175755
A:Accession: H81070
A:Status: preliminary
A:Status: type: DNA
A:Molecule type: DNA
A:Residues: 1-737 <TET>
A:Cross-references: GB:AE002504; GB:AE002098; MID:g7226785; PIDN:AAF41896.1; PID:g7222
A:Experimental source: serogroup B, strain MC58
C:Genetics:
A:Gene: NMB1541

```

[illegible][illegible]

C:Species: *Enterococcus faecalis*
C:Date: 30-Jun-1992 #sequence_revision 30-Jun-1992 #text_change 15-Oct-1999
C:Accession: G41662
R:Kao, S.M.; Olmsted, S.B.; Viksnins, A.S.; Gallo, J.C.; Dunny, G.M.
J. Bacteriol. 173, 7650-7664, 1991
A:Title: Molecular and genetic analysis of a region of plasmid pCF10 containing enterococcal faecalins.
A:Reference number: A41662; MUID: 92041679
A:Status: preliminary

GenCore version 5.1.1.3
Copyright (c) 1993 - 2002 Compugen Ltd.

OM protein - protein search, using sw model

Run on: October 29, 2002, 09:28:27 ; Search time 9.49256 Seconds
(without alignments)
784.805 Million cell updates/sec

Title: US-08-325-278B-1

Perfect score: 1565

Sequence: 1 AVENKETPTPTDSEEV.....GGYTINIRFAGKKVDEKPEE 305

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 231628 seqs, 24425594 residues

Total number of hits satisfying chosen parameters: 231628

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Issued_Patents_AA.*

- 1: /cgn2.6/ptodata/1/iaa/5A_COMB.pep.*
- 2: /cgn2.6/ptodata/1/iaa/5B_COMB.pep.*
- 3: /cgn2.6/ptodata/1/iaa/6A_COMB.pep.*
- 4: /cgn2.6/ptodata/1/iaa/6B_COMB.pep.*
- 5: /cgn2.6/ptodata/1/iaa/PTUS_COMB.pep.*
- 6: /cgn2.6/ptodata/1/iaa/backfiles1.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1565	100.0	305	2	US-08-795-475-1
2	1565	100.0	434	2	US-08-795-475-3
3	1226.5	78.4	1027	4	US-08-446-137B-2
4	1216	77.7	291	4	US-08-446-137B-4
5	618	39.5	342	3	US-08-828-741B-6
6	618	39.5	342	4	US-09-160-567-6
7	615	39.3	178	3	US-08-828-741B-13
8	615	39.3	178	4	US-09-160-567-13
9	615	39.3	198	3	US-08-828-741B-8
10	615	39.3	198	4	US-09-160-567-8
11	614	39.2	495	3	US-08-828-741B-4
12	614	39.2	495	4	US-09-160-567-4
13	613	39.2	182	3	US-08-828-741B-2
14	613	39.2	182	4	US-09-160-567-2
15	334	21.3	75	4	US-08-446-137B-8
16	315.5	20.2	71	4	US-08-446-137B-6
17	306	19.6	74	4	US-08-446-137B-7
18	303	19.4	71	4	US-08-446-137B-5
19	167	10.7	664	3	US-08-669-408B-2
20	151.5	9.7	1183	2	US-08-447-031A-2
21	128.5	8.2	564	4	US-09-308-022-6
22	127.5	8.1	631	4	US-08-847-065-25
23	123	7.9	936	5	PCR-US93-05944-2
24	121.5	7.8	522	4	US-08-961-083-120
25	121.5	7.8	1040	4	US-08-961-083-118
26	116.5	7.4	1912	1	US-08-409-995-4
27	116.5	7.4	1912	3	US-08-685-467-4

28	116.5	7.4	2353	4	US-09-377-155-33	Sequence 33, Appl
29	116.5	7.4	2353	4	US-08-913-942-4	Sequence 4, Appli
30	116.5	7.4	2353	4	US-09-669-974-33	Sequence 33, Appl
31	116.5	7.4	2354	4	US-09-268-347-47	Sequence 47, Appl
32	116	7.4	1104	4	US-09-268-347-28	Sequence 28, Appl
33	116	7.4	1104	4	US-09-268-347-34	Sequence 34, Appl
34	115	7.3	272	3	US-08-441-857-4	Sequence 4, Appli
35	115	7.3	272	3	US-08-193-159-4	Sequence 4, Appli
36	115	7.3	283	4	US-09-367-012-1	Sequence 1, Appli
37	115	7.3	337	2	US-08-441-857-8	Sequence 8, Appli
38	115	7.3	337	3	US-08-193-159-8	Sequence 8, Appli
39	115	7.3	365	3	US-08-483-577A-156	Sequence 156, App
40	115	7.3	365	4	US-08-897-438-156	Sequence 156, App
41	115	7.3	404	3	US-08-483-577A-155	Sequence 155, App
42	115	7.3	404	4	US-08-897-438-155	Sequence 155, App
43	115	7.3	411	3	US-08-483-577A-154	Sequence 154, App
44	115	7.3	411	4	US-08-897-438-154	Sequence 154, App
45	115	7.3	417	3	US-08-483-577A-153	Sequence 153, App

ALIGNMENTS

RESULT 1
US-08-795-475-1
; Sequence 1, Application US/08795475
; Patent No. 5965390
; GENERAL INFORMATION:
; APPLICANT: Bjvrck, Lars
; TITLE OF INVENTION: PROTEIN L AND HYBRID PROTEINS THEREOF
; NUMBER OF SEQUENCES: 14
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: SEED AND BERRY LLP
; STREET: 6300 Columbia Center, 701 Fifth Avenue
; CITY: Seattle
; STATE: Washington
; COUNTRY: USA
; ZIP: 98104-7092
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/795,475
; FILING DATE: 11-FEB-1997
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Mcmasters, David D.
; REGISTRATION NUMBER: 33,963
; REFERENCE/DOCKET NUMBER: 100084.402D1
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (206) 622-4900
; TELEFAX: (206) 682-6031
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 305 amino acids
; TYPE: amino acid
; STRANDEDNESS: unknown
; TOPOLOGY: unknown
; MOLECULE TYPE: protein
; HYPOTHETICAL: NO
; ORIGINAL SOURCE:
; ORGANISM: Escherichia coli LE392/pHDL, DSM 7054
US-08-795-475-1

DNAdiomed
seq2

Query Match 100.0%; Score 1565; DB 2; Length 305;
Best Local Similarity 100.0%; Pred. No. 1.6e-126;
Matches 305; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 AVENKETPTPTDSEEVITKANLIFANGSTQAFKCTFEKATSEAYAYADTLKKDN 60
|||||

Db 276 MERKSEKETPE--PEEVTIKANLIFADGSTQNAEFGTFAKAVSDAYAYADALKKNDG 333
QY 62 EYTVADVADKGYTLNFKAGKEKTPPEEVTIKANLIYADGKTQTAEFGTPEEATAEA 121
Db 334 EYTVADVADKGLTNFKAGKEKTPPEEVTIKVNLIFADGKTQTAEFGTPEEATAKA 393
QY 122 YRYADALKKONGEYTVADVADKGYTLNFKAGKE--KTPPEEKEEVTIKANLIYADGKTQT 179
Db 394 YAVADLLAKENGEYTVADLEDGGNTINIKFAGKETPETPEEKEEVTIKVNLIFADGKIQT 453
QY 180 AEFKGTPEEATAEAYRYADLLAKENGYTVADVADKGYTLNFKAGKE--KTPPEEKEEVT 237
Db 454 AEFKGTPEEATAEAYRYADLLAKENGYTVADLEDGGNTINIKFAGKETPETPEEKEEVT 513
QY 238 IKANLIYADGKTQTAEFGTFAEATAEAYRYADLLAKENGYTVADLEDGGVTTNIRFAGK 297
Db 514 IKVNLIFADGKTQTAEFGTFAEATAEAYRYADLLAKVNGEYTVADLEDGGVTTNIRFAGK 573
QY 298 KVDEKPEE 305
Db 574 ---EQPGE 578

RESULT 4

US-08-446-137B-4
; Sequence 4, Application US/08446137B
; Patent No. 6162903
; GENERAL INFORMATION:
; APPLICANT: Townen, Angus R.
; APPLICANT: Atkinson, Anthony
; APPLICANT: Murphy, Jonathan P.
; APPLICANT: Laurence, Oliver S.
; APPLICANT: Dugleby, Clive J.
; TITLE OF INVENTION: IMMUNOGLOBULIN BINDING PROTEINS DERIVED
; TITLE OF INVENTION: FROM L PROTEIN AND THEIR USES
; NUMBER OF SEQUENCES: 12
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: SEED and BERRY LLP
; STREET: 6300 Columbia Center, 701 Fifth Avenue
; CITY: Seattle
; STATE: Washington
; COUNTRY: USA
; ZIP: 98104
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/446,137B
; FILING DATE: 22-MAY-1995
; CLASSIFICATION: 514
; ATTORNEY/AGENT INFORMATION:
; NAME: McMaisters, David D.
; REGISTRATION NUMBER: 33,963
; REFERENCE/DOCKET NUMBER: 100084.406
; TELEPHONE: (206) 622-4900
; TELEFAX: (206) 682-6031
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 291 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-446-137B-4

Query Match 77.7%; Score 1216; DB 4; Length 291;
Best Local Similarity 82.5%; Pred. No. 1.1e-96;
Matches 241; Conservative 20; Mismatches 25; Indels 6; Gaps 3;
QY 10 ETPETDSEEEVTIKANLIFANGSTQTAEFKGTFAKATSEAYAYADTLKNDGGEYTVADVAD 69
||||| ||||||||| ||||| || ||||| |||||||||

Db 2 ETPE--PEEVTIKANLIFADGSTQNAEFGTFAKAVSDAYAYADALKKONGEYTVADVAD 59
QY 70 KGYTLNFKAGKEKTPPEEKEEVTIKANLIYADGKTQTAEFGTPEEATAEAYRYADALK 129
Db 60 KGYTLNFKAGKEKTPPEEKEEVTIKVNLIFADGKTQTAEFGTPEEATAKAYAYADLLA 119
QY 130 KONGEYTVADVADKGYTLNFKAGKE--KTPPEEKEEVTIKANLIYADGKTQTAEFGTPE 187
Db 120 KENGEYTVADLEDGGNTINIKFAGKETPETPEEKEEVTIKVNLIFADGKIQTAEFGTPE 179
QY 188 EATAEAYRYADLLAKENGYTVADVADKGYTLNFKAGKE--KTPPEEKEEVTIKANLIYA 245
Db 180 EATAKAYAYANLAKENGEYTVADLEDGGNTINIKFAGKETPETPEEKEEVTIKVNLIFA 239
QY 246 DGTQTAEFGTFAEATAEAYRYADLLAKENGYTVADLEDGGVTTNIRFAGK 297
Db 240 DGTQTAEFGTFAEATAEAYRYADLLAKVNGEYTVADLEDGGVTTNIRFAGK 291

RESULT 5

US-08-828-741B-6
; Sequence 6, Application US/08828741B
; Patent No. 6043069
; GENERAL INFORMATION:
; APPLICANT: Koentgen, Frank
; APPLICANT: Suess, Gabriele M.
; APPLICANT: Tarlinton, David M.
; APPLICANT: Treutlein, Herbert R.
; TITLE OF INVENTION: CATALYTIC ANTIBODIES AND A METHOD OF
; NUMBER OF SEQUENCES: 14
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: SCULLY, SCOTT, MURPHY & PRESSER
; STREET: 400 Garden City Plaza
; CITY: Garden City
; STATE: New York
; COUNTRY: United States of America
; ZIP: 11530
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/828,741B
; FILING DATE: 26-MAR-1997
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Digiglio, Frank S.
; REGISTRATION NUMBER: 31,346
; REFERENCE/DOCKET NUMBER: 10591
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (516) 742-4343
; TELEFAX: (516) 742-4366
; TELETYPE: 230 901 SANS UR
; INFORMATION FOR SEQ ID NO: 6:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 342 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-828-741B-6

Query Match 39.5%; Score 618; DB 3; Length 342;
Best Local Similarity 75.9%; Pred. No. 1.9e-45;
Matches 126; Conservative 5; Mismatches 15; Indels 20; Gaps 1;
QY 16 SEEEVTIKANLIFANGSTQTAEFKGTFAKATSEAYAYADTLKNDGGEYTVADVADKGYTLN 75
Db 174 SAEVTIKANLIFANGSTQTAEFKGTFAKATSEAYAYADTLKNDGGEYTVADVADKGYTLN 233
QY 76 IKFAGKEKTPPEE-----PKEEVTIKANLIYADGKTQTAEFGTPE 115
||||| | ||||||| :| |||||||

NUMBER OF SEQUENCES: 14
CORRESPONDENCE ADDRESS:
ADDRESSEE: SCULLY, SCOTT, MURPHY & PRESSER

NUMBER OF SEQUENCES: 14
CORRESPONDENCE ADDRESS:
ADDRESSEE: SCULLY, SCOTT, MURPHY & PRESSER

NUMBER OF SEQUENCES: 14
CORRESPONDENCE ADDRESS:
ADDRESSEE: SCULLY, SCOTT, MURPHY & PRESSER

STREET: 400 Garden City Plaza
CITY: Garden City
STATE: New York
COUNTRY: United States of America
ZIP: 11530
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/160,567
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/828,741
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: Digiglio, Frank S.
REGISTRATION NUMBER: 31,346
REFERENCE/DOCKET NUMBER: 10591
TELECOMMUNICATION INFORMATION:
TELEPHONE: (516) 742-4343
TELEFAX: (516) 742-4366
TELEX: 230 901 SANS UR
INFORMATION FOR SEQ ID NO: 13:
SEQUENCE CHARACTERISTICS:
LENGTH: 178 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-09-160-567-13

Query Match 39.3%; Score 615; DB 4; Length 178;
Best Local Similarity 76.2%; Pred. No. 1.4e-45;
Matches 125; Conservative 5; Mismatches 14; Indels 20; Gaps 1;

QY 18 EEVTKANLIFANGSTQTAFFKGTSEAYAYADTLKKDNGEYTVADVADKGYTLNIK 77
Db 12 EEVTKANLIFANGSTQTAFFKGTSEAYAYADTLKKDNGEYTVADVADKGYTLNIK 71
QY 78 FAGKEKTPPE-----PKKEVTIKANLIYADGKTQTAFFKGTFFEA 117
Db 72 FAGKEATNRNTDGTGYILQINRWGLTSAEEVTIKANLIFANGSTQTAFFKGTFFEA 131
QY 118 TABAYRYADALKKDNGEYTVADVADKGYTLNIKAGKEKTPPEPK 161
Db 132 TSEAYAYADTLKKDNGEYTVADVADKGYTLNIKAGKESAWRHPQ 175

RESULT 9
US-08-828-741B-8
Sequence 8, Application US/08828741B
Patent No. 6043069
GENERAL INFORMATION:
APPLICANT: Koentgen, Frank
APPLICANT: Suess, Gabriele M.
APPLICANT: Tarlinton, David M.
APPLICANT: Treutlein, Herbert R.
TITLE OF INVENTION: CATALYTIC ANTIBODIES AND A METHOD OF
OPERATING SYSTEM: PRODUCING SAME
NUMBER OF SEQUENCES: 14
CORRESPONDENCE ADDRESS:
ADDRESSEE: SCULLY, SCOTT, MURPHY & PRESSER
STREET: 400 Garden City Plaza
CITY: Garden City
STATE: New York
COUNTRY: United States of America
ZIP: 11530
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/828,741B
FILING DATE: 26-MAR-1997
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Digiglio, Frank S.
REGISTRATION NUMBER: 31,346
REFERENCE/DOCKET NUMBER: 10591
TELECOMMUNICATION INFORMATION:
TELEPHONE: (516) 742-4343
TELEFAX: (516) 742-4366
TELEX: 230 901 SANS UR
INFORMATION FOR SEQ ID NO: 8:
SEQUENCE CHARACTERISTICS:
LENGTH: 198 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-828-741B-8

Query Match 39.3%; Score 615; DB 3; Length 198;
Best Local Similarity 76.2%; Pred. No. 1.6e-45;
Matches 125; Conservative 5; Mismatches 14; Indels 20; Gaps 1;

QY 18 EEVTKANLIFANGSTQTAFFKGTSEAYAYADTLKKDNGEYTVADVADKGYTLNIK 77
Db 32 EEVTKANLIFANGSTQTAFFKGTSEAYAYADTLKKDNGEYTVADVADKGYTLNIK 91
QY 78 FAGKEKTPPE-----PKKEVTIKANLIYADGKTQTAFFKGTFFEA 117
Db 92 FAGKEATNRNTDGTGYILQINRWGLTSAEEVTIKANLIFANGSTQTAFFKGTFFEA 151
QY 118 TABAYRYADALKKDNGEYTVADVADKGYTLNIKAGKEKTPPEPK 161
Db 152 TSEAYAYADTLKKDNGEYTVADVADKGYTLNIKAGKESAWRHPQ 195

RESULT 10
US-09-160-567-8
Sequence 8, Application US/09160567
Patent No. 6326179
GENERAL INFORMATION:
APPLICANT: Koentgen, Frank
APPLICANT: Suess, Gabriele M.
APPLICANT: Tarlinton, David M.
APPLICANT: Treutlein, Herbert R.
TITLE OF INVENTION: CATALYTIC ANTIBODIES AND A METHOD OF
OPERATING SYSTEM: PRODUCING SAME
NUMBER OF SEQUENCES: 14
CORRESPONDENCE ADDRESS:
ADDRESSEE: SCULLY, SCOTT, MURPHY & PRESSER
STREET: 400 Garden City Plaza
CITY: Garden City
STATE: New York
COUNTRY: United States of America
ZIP: 11530
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/160,567
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/828,741
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: Digiglio, Frank S.
REGISTRATION NUMBER: 31,346
REFERENCE/DOCKET NUMBER: 10591

```
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (516) 742-4343
; TELEFAX: (516) 742-4366
; TELEX: 230 901 SANS UR
; INFORMATION FOR SEQ ID NO: 8:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 198 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-09-160-567-8

Query Match      39.3%; Score 615; DB 4; Length 198;
Best Local Similarity 76.2%; Pred. No. 1.6e-45;
Matches 125; Conservative 5; Mismatches 14; Indels 20; Gaps 1;

QY 18 EEEVTKANLIFANGSTQTAFFKGTFFKATSEAYAYADTLKKDNGEYTVDVADKGYTLN 77
DB 32 EEEVTKANLIFANGSTQTAFFKGTFFKATSEAYAYADTLKKDNGEYTVDVADKGYTLN 91
QY 78 FAKGKEKTPPEE-----PKEEVTKANLIYADGKTOTAEFKGTFFEA 117
DB 92 FAKGKEATNRNTDGDYGILOINSRWGGLTSAEEVTKANLIFANGSTQTAFFKGTFFEA 151
QY 118 TAPAYRYADALKKDNGEYTVDVADKGYTLNFKAGKEKTPPEEK 161
DB 152 TSEAYAYADTLKKDNGEYTVDVADKGYTLNFKAGKESAWRHQ 195

RESULT 11
US-08-828-741B-4
; Sequence 4, Application US/08828741B
; Patent No. 6043069
; GENERAL INFORMATION:
; APPLICANT: Koentgen, Frank
; APPLICANT: Suess, Gabriele M.
; APPLICANT: Tarlington, David M.
; APPLICANT: Treutlein, Herbert R.
; TITLE OF INVENTION: CATALYTIC ANTIBODIES AND A METHOD OF
; TITLE OF INVENTION: PRODUCING SAME
; NUMBER OF SEQUENCES: 14
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: SCULLY, SCOTT, MURPHY & PRESSER
; STREET: 400 Garden City Plaza
; CITY: Garden City
; STATE: New York
; COUNTRY: United States of America
; ZIP: 11530
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; FILING DATE: 26-MAR-1997
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Digiglio, Frank S.
; REGISTRATION NUMBER: 31,346
; REFERENCE/DOCKET NUMBER: 10591
; TELEPHONE: (516) 742-4343
; TELEFAX: (516) 742-4366
; TELEX: 230 901 SANS UR
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 495 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-828-741B-4

Query Match      39.3%; Score 615; DB 4; Length 198;
Best Local Similarity 76.2%; Pred. No. 1.6e-45;
Matches 125; Conservative 5; Mismatches 14; Indels 20; Gaps 1;

QY 18 EEEVTKANLIFANGSTQTAFFKGTFFKATSEAYAYADTLKKDNGEYTVDVADKGYTLN 77
DB 32 EEEVTKANLIFANGSTQTAFFKGTFFKATSEAYAYADTLKKDNGEYTVDVADKGYTLN 91
QY 78 FAKGKEKTPPEE-----PKEEVTKANLIYADGKTOTAEFKGTFFEA 117
DB 92 FAKGKEATNRNTDGDYGILOINSRWGGLTSAEEVTKANLIFANGSTQTAFFKGTFFEA 151
QY 118 TAPAYRYADALKKDNGEYTVDVADKGYTLNFKAGKEKTPPEEK 161
DB 152 TSEAYAYADTLKKDNGEYTVDVADKGYTLNFKAGKESAWRHQ 195

RESULT 12
US-09-160-567-4
; Sequence 4, Application US/09160567
; Patent No. 6326179
; GENERAL INFORMATION:
; APPLICANT: Koentgen, Frank
; APPLICANT: Suess, Gabriele M.
; APPLICANT: Tarlington, David M.
; APPLICANT: Treutlein, Herbert R.
; TITLE OF INVENTION: CATALYTIC ANTIBODIES AND A METHOD OF
; TITLE OF INVENTION: PRODUCING SAME
; NUMBER OF SEQUENCES: 14
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: SCULLY, SCOTT, MURPHY & PRESSER
; STREET: 400 Garden City Plaza
; CITY: Garden City
; STATE: New York
; COUNTRY: United States of America
; ZIP: 11530
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION NUMBER: US/09/160,567
; APPLICATION NUMBER: 08/828,741
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Digiglio, Frank S.
; REGISTRATION NUMBER: 31,346
; REFERENCE/DOCKET NUMBER: 10591
; TELEPHONE: (516) 742-4343
; TELEFAX: (516) 742-4366
; TELEX: 230 901 SANS UR
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 495 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-09-160-567-4

Query Match      39.2%; Score 614; DB 4; Length 495;
Best Local Similarity 78.6%; Pred. No. 6.8e-45;
Matches 125; Conservative 4; Mismatches 10; Indels 20; Gaps 1;

QY 16 SEEEVTKANLIFANGSTQTAFFKGTFFKATSEAYAYADTLKKDNGEYTVDVADKGYTLN 75
DB 174 SAEVTKANLIFANGSTQTAFFKGTFFKATSEAYAYADTLKKDNGEYTVDVADKGYTLN 233
QY 76 IKFAGKEKTPPEE-----PKEEVTKANLIYADGKTOTAEFKGTFFE 115
DB 234 IKFAGKEATNRNTDGDYGILOINSRWGGLTSAEEVTKANLIFANGSTQTAFFKGTFFE 293
QY 116 EATAPAYRYADALKKDNGEYTVDVADKGYTLNFKAGKE 154
DB 294 KATSEAYAYADTLKKDNGEYTVDVADKGYTLNFKAGKE 332

Query Match      39.2%; Score 614; DB 4; Length 495;
Best Local Similarity 78.6%; Pred. No. 6.8e-45;
Matches 125; Conservative 4; Mismatches 10; Indels 20; Gaps 1;

QY 16 SEEEVTKANLIFANGSTQTAFFKGTFFKATSEAYAYADTLKKDNGEYTVDVADKGYTLN 75
DB 174 SAEVTKANLIFANGSTQTAFFKGTFFKATSEAYAYADTLKKDNGEYTVDVADKGYTLN 233
QY 76 IKFAGKEKTPPEE-----PKEEVTKANLIYADGKTOTAEFKGTFFE 115
```



```

||||| I
Db 234 IKFAGKEATNRNTDGSYDYLQINRWGGLTSAEEVTIKANLIFANGSTQTAEFKGTFF 293
QY 116 EATAEAYRYADALKKNGEYTVDVADKGYTLINIKFAGKE 154
:||||| I
Db 294 KATSEAYAYADTLKKNGEYTVDVADKGYTLINIKFAGKE 332

RESULT 13
US-08-828-741B-2
; Sequence 2, Application US/08828741B
; Patent No. 6043069
; GENERAL INFORMATION:
; APPLICANT: Koentgen, Frank
; APPLICANT: Suess, Gabriele M.
; APPLICANT: Tarlinton, David M.
; APPLICANT: Treutlein, Herbert R.
; TITLE OF INVENTION: CATALYTIC ANTIBODIES AND A METHOD OF
; NUMBER OF SEQUENCES: 14
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: SCULLY, SCOTT, MURPHY & PRESSER
; STREET: 400 Garden City Plaza
; CITY: Garden City
; STATE: New York
; COUNTRY: United States of America
; ZIP: 11530
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/828, 741B
; FILING DATE: 26-MAR-1997
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Digiglio, Frank S.
; REGISTRATION NUMBER: 31,346
; REFERENCE/DOCKET NUMBER: 10591
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (516) 742-4366
; TELEFAX: (516) 742-4366
; TELEX: 230 901 SANS UR
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 182 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-828-741B-2

Query Match 39.2%; Score 613; DB 3; Length 182;
Best Local Similarity 78.5%; Pred. No. 2.le-45;
Matches 124; Conservative 5; Mismatches 9; Indels 20; Gaps 1;

QY 15 DSEEVTKANLIFANGSTQTAEFKGTSEAYAYADTLKKNGEYTVDVADKGYTL 74
I:||||| I
Db 25 DNTBEVTIKANLIFANGSTQTAEFKGTSEAYAYADTLKKNGEYTVDVADKGYTL 84

QY 75 NIKFAGKEKTPPEP-----KEEVTIKANLIYADGKTQTAEFKGTFF 114
||||| I
Db 85 NIKFAGKEATNRNTDGSYDYLQINRWGGLTLKEEVTIKANLIFANGSTQTAEFKGTFF 144

QY 115 EATAEAYRYADALKKNGEYTVDVADKGYTLINIKFAG 152
I:||||| I
Db 145 EKATSEAYAYADTLKKNGEYTVDVADKGYTLINIKFAG 182

RESULT 14
US-09-160-567-2
; Sequence 2, Application US/09160567
; Patent No. 6326179

```

```

; GENERAL INFORMATION:
; APPLICANT: Koentgen, Frank
; APPLICANT: Suess, Gabriele M.
; APPLICANT: Tarlinton, David M.
; APPLICANT: Treutlein, Herbert R.
; TITLE OF INVENTION: CATALYTIC ANTIBODIES AND A METHOD OF
; NUMBER OF SEQUENCES: 14
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: SCULLY, SCOTT, MURPHY & PRESSER
; STREET: 400 Garden City Plaza
; CITY: Garden City
; STATE: New York
; COUNTRY: United States of America
; ZIP: 11530
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/160,567
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/828,741
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Digiglio, Frank S.
; REGISTRATION NUMBER: 31,346
; REFERENCE/DOCKET NUMBER: 10591
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (516) 742-4343
; TELEFAX: (516) 742-4366
; TELEX: 230 901 SANS UR
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 182 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-09-160-567-2

Query Match 39.2%; Score 613; DB 4; Length 182;
Best Local Similarity 78.5%; Pred. No. 2.le-45;
Matches 124; Conservative 5; Mismatches 9; Indels 20; Gaps 1;

QY 15 DSEEVTKANLIFANGSTQTAEFKGTSEAYAYADTLKKNGEYTVDVADKGYTL 74
I:||||| I
Db 25 DNTBEVTIKANLIFANGSTQTAEFKGTSEAYAYADTLKKNGEYTVDVADKGYTL 84

QY 75 NIKFAGKEKTPPEP-----KEEVTIKANLIYADGKTQTAEFKGTFF 114
||||| I
Db 85 NIKFAGKEATNRNTDGSYDYLQINRWGGLTLKEEVTIKANLIFANGSTQTAEFKGTFF 144

QY 115 EATAEAYRYADALKKNGEYTVDVADKGYTLINIKFAG 152
I:||||| I
Db 145 EKATSEAYAYADTLKKNGEYTVDVADKGYTLINIKFAG 182

RESULT 15
US-08-446-137B-8
; Sequence 8, Application US/08446137B
; Patent No. 6162903
; GENERAL INFORMATION:
; APPLICANT: Trowen, Angus R.
; APPLICANT: Atkinson, Anthony
; APPLICANT: Murphy, Jonathan P.
; APPLICANT: Laurence, Oliver S.
; APPLICANT: Duggleby, Clive J.
; TITLE OF INVENTION: IMMUNOGLOBULIN BINDING PROTEINS DERIVED
; FROM L PROTEIN AND THEIR USES
; NUMBER OF SEQUENCES: 12

```

CORRESPONDENCE ADDRESS:
 ADDRESSEE: SEED AND BERRY LLP
 STREET: 6300 Columbia Center, 701 Fifth Avenue
 CITY: Seattle
 STATE: Washington
 COUNTRY: USA
 ZIP: 98104
 COMPUTER READABLE FORM:
 MEDIUM TYPE: Floppy disk
 COMPUTER: IBM PC compatible
 OPERATING SYSTEM: PC-DOS/MS-DOS
 SOFTWARE: Patent In Review #1.0, Version #1.30
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/08/446.137B
 FILING DATE: 22-MAY-1995
 CLASSIFICATION: 514
 ATTORNEY/AGENT INFORMATION:
 NAME: Mcmasters, David D.
 REGISTRATION NUMBER: 33,963
 REFERENCE/DOCKET NUMBER: 100084.406
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: (206) 622-4900
 TELEFAX: (206) 682-6031
 INFORMATION FOR SEQ ID NO: 8:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 75 amino acids
 TYPE: amino acid
 STRANDEDNESS:
 TOPOLOGY: linear
 US-08-446-137B-8

```

Query Match      21.3%; Score 334; DB 4; Length 75;
Best-Local Similarity 90.1%; Pred. No. 4.5e-23;
Matches 64; Conservative 4; Mismatches 3; Indels 0; Gaps 0;

QY 227 KTPPEPKKEVTIKANLIYADGKTTQTAEPKGTFAZATAEAYRYADLLAKENSKYTDLEDG 286
      :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 5 ETPEPKKEVTKVNLIFADGKTQTAEPKGTFEZATAEAYRYADLLAKVNGEYTDLEDG 64

QY 287 GYTINIRFAGK 297
      |||||:|||||
Db 65 GYTINIRFAGK 75

```

Search completed: October 29, 2002, 09:31:54
Job time : 11.4926 secs

GenCore version 5.1.3
Copyright (c) 1993 - 2002 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: October 29, 2002, 09:25:06 ; Search time 23.9378 Seconds
(without alignments)
1415.233 Million cell updates/sec

Title: US-08-325-278b-1

Perfect score: 1565

Sequence: 1 AVENKEETPETDSEEV.....GGYTINIRFAGKKVDEKPE 305

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 747574 seqs, 111073796 residues

Total number of hits satisfying chosen parameters: 747574

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : A_Geneseq_032802.*
1: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1980.DAT.*
2: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1981.DAT.*
3: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1982.DAT.*
4: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1983.DAT.*
5: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1984.DAT.*
6: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1985.DAT.*
7: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1986.DAT.*
8: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1987.DAT.*
9: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1988.DAT.*
10: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1989.DAT.*
11: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1990.DAT.*
12: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1991.DAT.*
13: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1992.DAT.*
14: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1993.DAT.*
15: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1994.DAT.*
16: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1995.DAT.*
17: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1996.DAT.*
18: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1997.DAT.*
19: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1998.DAT.*
20: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1999.DAT.*
21: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA2000.DAT.*
22: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA2001.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1565	100.0	305	14	AA42993
2	1565	100.0	434	14	AA42994
3	1544.5	98.7	22	22	AA31372
4	1366	87.3	367	21	AA10432
5	1226.5	78.4	1027	14	AA42203
6	1226.5	78.4	1027	14	AA43699
7	1216	77.7	291	14	AA42204
8	618	39.5	342	18	AA32482
9	618	39.5	342	20	AA706909
10	615	39.3	178	18	AAW32486
11	615	39.3	178	20	AAW06913

1999 date
- 2000 date

12	615	39.3	198	18	AAW32481	Growth factor LHL.
13	615	39.3	198	20	AAW06910	Amino acid sequenc
14	614	39.2	495	18	AAW32480	Growth factor CATA
15	614	39.2	495	20	AAW06908	CATAB-TEV aminoaci
16	613	39.2	182	18	AAW32479	Growth factor LHL
17	613	39.2	182	20	AAW06907	LHL growth factor
18	608	38.8	482	20	AAW06915	Amino acid sequenc
19	389	24.9	76	21	AAW82537	Peptostreptococcus
20	371	23.7	72	21	AAW82538	Peptostreptococcus
21	370	23.6	72	21	AAW82540	Peptostreptococcus
22	369	23.6	72	21	AAW82539	Peptostreptococcus
23	341	21.8	82	21	AAW82536	Ig light chain bin
24	334	21.3	75	21	AAW82544	Peptostreptococcus
25	334	21.3	82	21	AAW82545	PpL mutant protein
26	334	21.3	82	21	AAW82548	PpL mutant protein
27	332	21.2	82	21	AAW82546	PpL mutant protein
28	328	21.0	82	21	AAW82547	PpL mutant protein
29	315.5	20.2	71	21	AAW82542	Peptostreptococcus
30	306	19.6	74	21	AAW82543	Peptostreptococcus
31	303	19.4	71	21	AAW82541	Peptostreptococcus
32	167	10.7	664	16	AAW71929	S. dysgalactiae MI
33	154.5	9.9	1185	13	AAW22675	Collagen binding p
34	149	9.5	448	21	AAW77965	A. thaliana enviro
35	141.5	9.0	1638	20	AAW00138	Enterococcus faeca
36	141.5	9.0	1638	20	AAW00140	Enterococcus faeca
37	141.5	9.0	1638	20	AAW00142	Enterococcus faeca
38	131.5	8.4	414	20	AAW49150	Amino acid sequenc
39	131.5	8.4	1166	20	AAW08643	S. aureus SdrE pro
40	130.5	8.3	376	20	AAW49231	N-terminal truncat
41	130.5	8.3	376	20	AAW32103	Choline binding pr
42	130.5	8.3	413	20	AAW49247	N-terminal region
43	130.5	8.3	413	20	AAW32186	N-terminal choline
44	130.5	8.3	414	20	AAW49144	Amino acid sequenc
45	130.5	8.3	564	19	AAW62654	C3 binding protein

ALIGNMENTS

RESULT 1	AA42993	standard; Protein; 305 AA.
ID	AA42993	standard; Protein; 305 AA.
XX		
AC	AA42993	
XX		
XX		
DT	16-MAY-1994	(first entry)
XX		
DE	Immunoglobulin light chain binding protein (Protein L).	
DE	Immunoglobulin; light chain; binding; identification; purification;	
KW	separation.	
KW		
XX		
OS	E. coli LE392/pHDL, DSM 7054.	
XX		
FH	Key	Location/Qualifiers
FT	Domain	5..305
FT		/label= B1 immunoglobulin light chain binding
FT		domain.
FT	Domain	81..305
FT		/label= B2 immunoglobulin light chain binding
FT		domain.
FT	Domain	153..305
FT		/label= B3 immunoglobulin light chain binding
FT		domain.
FT	Domain	225..305
FT		/label= B4 immunoglobulin light chain binding
FT		domain.
FT	Domain	297..305
FT		/label= B5 immunoglobulin light chain binding
FT		domain.
XX		
PN	WO9323242-A.	

```

PD 11-NOV-1993.
XX
PF 28-APR-1993; 93WO-SE00375.
XX
PR 28-APR-1992; 92SE-0001331.
XX
XX (HIGH-) HIGHTECH RECEPTOR AB.
XX
XX Bjoerck L, Sjoerbring U;
XX WPI; 1993-368722/46.
XX DR N-PSDB; AAQ50452.
XX
XX New protein L binding light chains of all immunoglobulin classes
PT - for binding purifying and identifying immunoglobulin, also
PT related DNA, vectors and host cells
XX
PS Claim 1; Page 36; 71pp; English.
XX
XX The protein (Protein L) is capable of binding to immunoglobulin G
CC light chains. It is useful for binding, separating (purifying) and
CC identifying immunoglobulin and for removing immunoglobulin molecules
CC from serum. Hybrid proteins of the L protein can bind all human
CC immunoglobulin classes and many immunoglobulins from other species.
CC They are highly soluble and retain their binding activity at high
CC temperatures over a pH range of 3-10. They can be immobilised
CC without loss of activity.
XX
XX Sequence 305 AA;
XX
Query Match 100.0%; Score 1565; DB 14; Length 305;
Best Local Similarity 100.0%; Pred. No. 2.2e-110;
Matches 305; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 AVENKEETPETDSEEVTIKANLIFANGSTQTAFFKGTFEKATSEAYADTLKKN 60
DB 1 AVENKEETPETDSEEVTIKANLIFANGSTQTAFFKGTFEKATSEAYADTLKKN 60
QY 61 GEYTVADVADKGYTLNKFAGKEKTPPEPKKEVTIKANLIYADGKTQTAFFKGTFEATAE 120
DB 61 GEYTVADVADKGYTLNKFAGKEKTPPEPKKEVTIKANLIYADGKTQTAFFKGTFEATAE 120
QY 121 AYRYADALKKNGEYTVADVADKGYTLNKFAGKEKTPPEPKKEVTIKANLIYADGKTQTA 180
DB 121 AYRYADALKKNGEYTVADVADKGYTLNKFAGKEKTPPEPKKEVTIKANLIYADGKTQTA 180
QY 181 EFKGTFEATAEAYRYADLLAKENGYTVADVADKGYTLNKFAGKEKTPPEPKKEVTIKA 240
DB 181 EFKGTFEATAEAYRYADLLAKENGYTVADVADKGYTLNKFAGKEKTPPEPKKEVTIKA 240
QY 241 NLIYADGKTQTAFFKGTFAEATAEAYRYADLLAKENGYTADLEDGGYTNIRFAGKKVD 300
DB 241 NLIYADGKTQTAFFKGTFAEATAEAYRYADLLAKENGYTADLEDGGYTNIRFAGKKVD 300
QY 301 EKPEE 305
DB 301 EKPEE 305

```

RESULT 2

AAR42994
ID AAR42994 standard; Protein; 434 AA.

XX AAR42994;

XX 16-MAY-1994 (first entry)

XX Sequence encoding immunoglobulin light chain binding protein.

XX Immunoglobulin; light chain; binding; identification; purification;
XX separation; ss.

XX E. coli L392/pHDLG, DSM 7055.

```

XX
FH Location/Qualifiers
FT 5..305
FT /label= B1 immunoglobulin light chain binding
FT domain.
FT 81..305
FT /label= B2 immunoglobulin light chain binding
FT domain.
FT 153..305
FT /label= B3 immunoglobulin light chain binding
FT domain.
FT 225..305
FT /label= B4 immunoglobulin light chain binding
FT domain.
FT 297..305
FT /label= B5 immunoglobulin light chain binding
FT domain.
FT 309..434
FT /label= C1 immunoglobulin heavy chain binding
FT domain.
FT 364..434
FT /label= D intermediate immunoglobulin heavy
FT chain binding domain.
FT 379..434
FT /label= C2 immunoglobulin heavy chain binding
FT domain.
XX
XX WO9322342-A.
XX
XX 11-NOV-1993.
XX
XX 28-APR-1993; 93WO-SE00375.
XX
XX 28-APR-1992; 92SE-0001331.
XX
XX (HIGH-) HIGHTECH RECEPTOR AB.
XX
XX Bjoerck L, Sjoerbring U;
XX WPI; 1993-368722/46.
XX DR P-PSDB; AAR42994.
XX
XX New protein L binding light chains of all immunoglobulin classes
PT - for binding purifying and identifying immunoglobulin, also
PT related DNA, vectors and host cells
XX
XX Claim 6; Page 39-40; 71pp; English.
XX
XX Protein L (AAR42993) is capable of binding to immunoglobulin G light
CC chains. It is useful for binding, separating (purifying) and
CC identifying immunoglobulin and for removing immunoglobulin molecules
CC from serum. This is the coding sequence of one hybrid protein of
CC the L protein. The hybrid proteins can bind all human
CC immunoglobulin classes and many immunoglobulins from other species.
CC They are highly soluble and retain their binding activity at high
CC temperatures over a pH range of 3-10. They can be immobilised
CC without loss of activity.
XX
XX Sequence 434 AA;
XX
Query Match 100.0%; Score 1565; DB 14; Length 434;
Best Local Similarity 100.0%; Pred. No. 3.4e-110;
Matches 305; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 AVENKEETPETDSEEVTIKANLIFANGSTQTAFFKGTFEKATSEAYADTLKKN 60
DB 1 AVENKEETPETDSEEVTIKANLIFANGSTQTAFFKGTFEKATSEAYADTLKKN 60
QY 61 GEYTVADVADKGYTLNKFAGKEKTPPEPKKEVTIKANLIYADGKTQTAFFKGTFEATAE 120
DB 61 GEYTVADVADKGYTLNKFAGKEKTPPEPKKEVTIKANLIYADGKTQTAFFKGTFEATAE 120
QY 121 AYRYADALKKNGEYTVADVADKGYTLNKFAGKEKTPPEPKKEVTIKANLIYADGKTQTA 180

```

Db 121 AYRYADALKKDNGETYVDVADKGYTLNKFAGKEKTPPEPKKEEVTIKANLIYADGKTQTA 180
 QY 181 EFKGTPEEATAEAYRYADLLAKENGYTVDVADKGYTLNKFAGKEKTPPEPKKEEVTIKA 240
 Db 181 EFKGTPEEATAEAYRYADLLAKENGYTVDVADKGYTLNKFAGKEKTPPEPKKEEVTIKA 240
 QY 241 NLIYADGKTQTAEFKGTFAEATAEAYRYADLLAKENGYTADLEGGYTNIRFAGKKVD 300
 Db 241 NLIYADGKTQTAEFKGTFAEATAEAYRYADLLAKENGYTADLEGGYTNIRFAGKKVD 300
 QY 301 EKPEE 305
 Db 301 EKPEE 305
 RESULT 3
 AAB31372
 ID AAB31372 standard; Protein: 467 AA.
 AC AAB31372;
 XX
 DT 20-APR-2001 (first entry)
 XX
 DE Amino acid sequence of protein L/CBD cex/ER retaining peptide fusion.
 KW Protein production; food processing; protein antibiotic; feed enzyme;
 KW protein L: CBD cex protein; cell signal peptide.
 XX
 OS Synthetic.
 XX
 PN W0200077174-A1.
 XX
 PD 21-DEC-2000.
 XX
 PF 07-JUN-2000; 2000WO-IL00330.
 XX
 PR 10-JUN-1999; 99US-0329234.
 XX
 PA (CBDT-) CBD TECHNOLOGIES LTD.
 PA (YISS) YISSUM RES DEV CO HEBREW UNIV JERUSALEM.
 XX
 PI Shani Z, Shoseyov O;
 XX
 WPI: 2001-112219/12.
 DR N-PSDB; AAF24730.
 XX
 XX
 PT Expressing and isolating recombinant protein in a plant, useful for
 PT producing large quantities of recombinant proteins, by expressing a
 PT fusion protein including a cellulose binding peptide fused to a
 PT recombinant protein -
 XX
 PS Example: Fig 2a; 87pp; English.
 XX
 CC The specification describes a method for expressing and isolating
 CC a recombinant protein in a plant. The method comprising expressing a
 CC fusion protein including the recombinant protein and a cellulose
 CC binding peptide fused to it, where the fusion protein is
 CC compartmentalised and sequestered within plant cells, plant derived
 CC tissue or cultured plant cells. The method is useful for obtaining large
 CC quantities of the recombinant proteins and protein products in a simple
 CC and cost-effective manner. Recombinant proteins may be used commercially,
 CC such as in the food processing industry, e.g. glucoamylases and glucose
 CC isomerases are used for converting starch to high fructose corn syrup,
 CC proteinases for the hydrolysis of high molecular weight proteins and in
 CC manufacturing leather or alcoholic beverages, pectinesterases for
 CC pectin hydrolysis in food industry, lipases for cleaving ester linkage
 CC in triglycerides, and for effluent treatment. The recombinant proteins
 CC may further be used to produce protein antibiotics, which can be used
 CC in healing processes, and to produce animal feed enzymes. The present
 CC sequence represents a fusion protein of the invention, and comprises a
 CC fusion of a cell signal peptide, protein L, CBD cex, and an endoplasmic
 CC reticulum retaining peptide.

XX
 SQ Sequence 467 AA;
 Query Match 98.7%; Score 1544.5; DB 22; Length 467;
 Best Local Similarity 99.3%; Pred. No. 1.3e-108;
 Matches 303; Conservative 0; Mismatches 1; Indels 1; Gaps 1;
 QY 1 AVENKEETPTPTDSEEEVTIKANLIFANGSTQTAEFKGTPEKATSEAYAYADTLKKDN 60
 Db 40 AVENKEETPTPTDSEEEVTIKANLIFANGSTQTAEFKGTPEKATSEAYAYADTLKKDN 99
 QY 61 GEYTVADVADKGYTLNKFAGKEKTPPEPKKEEVTIKANLIYADGKTQTAEFKGTPEEATAE 120
 Db 100 GEYTVADVADKGYTLNKFAGKEKTPPEPKKEEVTIKANLIYADGKTQTAEFKGTPEEATAE 159
 QY 121 AYRYADALKKDNGETYVDVADKGYTLNKFAGKEKTPPEPKKEEVTIKANLIYADGKTQTA 180
 Db 160 AYRYADALKKDNGETYVDVADKGYTLNKFAGKEKTPPEPKKEEVTIKANLIYADGKTQTA 219
 QY 181 EFKGTPEEATAEAYRYADLL-ARENGKTYTVDVADKGYTLNKFAGKEKTPPEPKKEEVTIK 239
 Db 220 EFKGTPEEATAEAYRYADLLAAKENGKTYTVDVADKGYTLNKFAGKEKTPPEPKKEEVTIK 279
 QY 240 ANLIYADGKTQTAEFKGTFAEATAEAYRYADLLAKENGYTADLEGGYTNIRFAGKKV 299
 Db 280 ANLIYADGKTQTAEFKGTFAEATAEAYRYADLLAKENGYTADLEGGYTNIRFAGKKV 339
 QY 300 DEKPE 304
 Db 340 DEKPE 344
 RESULT 4
 AAB10432
 ID AAB10432 standard; Protein: 367 AA.
 AC AAB10432;
 XX
 DT 01-DEC-2000 (first entry)
 XX
 DE Expression vector pSEX1114 protein G.
 KW Expression vector; antibody binding protein; monoclonal antibody; Neo-R;
 KW B lymphocyte; myeloma cell; hybridoma; protein G; bla protein.
 XX
 OS Synthetic.
 XX
 PN DE19900635-A1.
 XX
 PD 13-JUL-2000.
 XX
 PF 11-JAN-1999; 99DE-1000635.
 XX
 PR 11-JAN-1999; 99DE-1000635.
 XX
 PA (DEKR-) DEUT KREBSFORSCHUNGSZENTRUM.
 XX
 PI Breitling F, Poustka A, Moldenhauer G;
 XX
 WPI: 2000-499832/45.
 DR N-PSDB; AAA71428.
 XX
 PT Selecting monoclonal antibodies, by expressing them on the surface of
 PT hybridomas attached to antibody-binding protein, then reaction with
 PT antibody library -
 XX
 PS Claim 16; Fig 1; 22pp; German.
 CC This invention describes a novel method for the selection of monoclonal
 CC antibodies (MAB) which comprises (i) fusing B lymphocytes with myeloma
 CC cells to produce antibody-producing hybridomas such that the antibodies
 CC are presented at the surface of the hybridomas by an antibody-binding
 CC protein (I); and (ii) binding the antibody to antigens (Ag). The

CC invention also describes antibody-binding proteins (I) that comprise a
 CC combination of the signal peptide of a murine immunoglobulin (Ig) kappa
 CC chain or a murine MHC (major histocompatibility complex) Class I k(k)
 CC molecule; an antibody-binding site of proteins A, G, L or Ig, and the
 CC transmembrane domain of PDGFR (platelet-derived growth factor receptor)
 CC or CD52. The method is used to select MAB with specificity for particular
 CC antigens. MAB can be selected without separate culture of hybridomas,
 CC and selection can be made against many antigens in a library, optionally
 CC on the basis of strength of affinity for a particular antigen. Complex
 CC mixtures of hybridomas can be used for selection, reducing the time and
 CC cost involved in MAB selection. This sequence represents the protein G,
 CC contained in the expression vector pSEX114 which contains the protein G,
 CC Neo-R and the bla protein described in the method of the invention.
 CC
 XX Sequence 367 AA;

Query Match 87.3%; Score 1366; DB 21; Length 367;
 Best Local Similarity 88.6%; Pred. No. 2.8e-95;
 Matches 264; Conservative 15; Mismatches 15; Indels 4; Gaps 1;
 QY 5 KEETPTPTDSEEVITKANLIFANGSTQTAFFKGTFFKATSEAYAYADTLKNGEY 64
 DB 25 KETPEEP---KEEVTIRANLIYADGKTQTAFFKGTFFKATSEAYAYADTLKNGEY 80
 QY 65 VDADKGYTLNIFAGKEKTPPEPKKEVTIKANLIYADGKTQTAFFKGTFFKATSEAYAY 124
 DB 81 VDADKGYTLNIFAGKEKTPPEPKKEVTIKANLIYADGKTQTAFFKGTFFKATSEAYAY 140
 QY 125 ADALKKNGEYTVDVADKGYTLNIFAGKEKTPPEPKKEVTIKANLIYADGKTQTAFFK 184
 DB 141 ADALKKNGEYTVDVADKGYTLNIFAGKEKTPPEPKKEVTIKANLIYADGKTQTAFFK 200
 QY 185 TFEATAEAYRYADLLAKENGKGYTVDVADKGYTLNIFAGKEKTPPEPKKEVTIKANLIY 244
 DB 201 TFEATAEAYRYADLLAKENGKGYTVDVADKGYTLNIFAGKEKTPPEPKKEVTIKANLIY 260
 QY 245 ADGKTQTAFFKGTFFKATSEAYAYADLLAKENGKGYTVADLEGGYTTINIRFAGK 302
 DB 261 ADGKTQTAFFKGTFFKATSEAYAYADLLAKENGKGYTVDVADKGYTLNIFAGK 318

RESULT 5
 AAR42203
 ID AAR42203 standard; Protein; 1027 AA.
 AC AAR42203;
 XX 18-MAY-1994 (first entry)
 DT Protein L.
 DE
 XX Peptide; immunoglobulin; binding; analysis; purification; ELISA;
 KW enzyme linked immunoabsorbant assay.
 XX
 OS Peptococcus magnus.

Key Location/Qualifiers
 FT Peptide 36..59
 FT /label= Signal sequence.
 FT Protein 60..968
 FT /label= Mature protein L.
 XX
 PN W09322439-A.
 XX
 PD 11-NOV-1993.
 XX
 PF 07-MAY-1993; 93WO-GB00950.
 XX
 PR 07-MAY-1992; 92GB-0009804.
 PR 24-DEC-1992; 92GB-0026928.
 XX
 PA (PUBL-) PUBLIC HEALTH LAB SERVICE BOARD.
 XX

PI Atkinson A, Duggleby CJ, Murphy JP, Trowern AR;
 XX WPI; 1993-368798/46.
 DR N-PSDB; AAQ50946.
 XX
 PT New immunoglobulin binding proteins derived from Protein L -
 PT cell walls
 XX
 PS Disclosure; Figure 1; 28pp; English.
 XX
 CC The synthetic immunoglobulin binding proteins derived from protein
 CC L correspond to the repeated sequences in protein L which bind
 CC immunoglobulin kappa light chains. They can be used in protein
 CC analysis, purification procedures and other biochemical processes e.
 CC g. ELISA. The synthetic molecules are of particular advantage if
 CC they are free of regions in protein L which exhibit albumin and cell
 CC wall binding.
 CC
 XX Sequence 1027 AA;

Query Match 78.4%; Score 1226.5; DB 14; Length 1027;
 Best Local Similarity 79.9%; Pred. No. 3.2e-84;
 Matches 246; Conservative 22; Mismatches 31; Indels 9; Gaps 4;
 QY 2 VENKEETPTDSEEVITKANLIFANGSTQTAFFKGTFFKATSEAYAYADTLKNG 61
 DB 276 MERKLEKETPE--PEEEVTIKANLIFADGSGTQNAEFGTFAKAVSDAYADALCKNG 333
 QY 62 EYTVADKGYTLNIFAGKEKTPPEPKKEVTIKANLIYADGKTQTAFFKGTFFKATSEAY 121
 DB 334 EYTVADKGYTLNIFAGKEKTPPEPKKEVTIKANLIYADGKTQTAFFKGTFFKATSEAY 393
 QY 122 YRYADALKKNGEYTVDVADKGYTLNIFAGKE--KTPPEKKEVTIKANLIYADGKT 179
 DB 394 YAYADLLAKENGKGYTVADLEGGYTTINIRFAGKTPPEPKKEVTIKANLIYADGKT 453
 QY 180 AEFKGTFFKATSEAYAYADLLAKENGKGYTVDVADKGYTLNIFAGKE--KTPPEKKEVT 237
 DB 454 AEFKGTFFKATSEAYAYADLLAKENGKGYTVADLEGGYTTINIRFAGKTPPEPKKEVT 513
 QY 238 IKANLIYADGKTQTAFFKGTFAEATAEAYRYADLLAKENGKGYTVADLEGGYTTINIRFAGK 297
 DB 514 IKVNLIFADGKTQTAFFKGTFFKATSEAYAYRYADLLAKENGKGYTVADLEGGYTTINIRFAGK 573
 QY 298 KVDEKPEE 305
 DB 574 ---EQPGE 578

RESULT 6
 AAR43699
 ID AAR43699 standard; Protein; 1027 AA.
 AC AAR43699;
 XX 18-MAY-1994 (first entry)
 DT Protein L.
 DE
 XX Peptide; immunoglobulin; binding; immobilisation; light chains;
 KW antibodies; diagnosis; pharmaceutical; ss.
 XX
 OS Peptococcus magnus.
 XX
 FH Key Location/Qualifiers
 FT Peptide 36..59
 FT /label= Signal sequence.
 FT Protein 60..968
 FT /label= Mature protein L.
 XX
 PN W09322438-A.
 XX

PD 02-OCT-1997.
 XX
 PF 26-MAR-1997; 97WO-AU00194.
 XX
 XX 27-FEB-1997; 97AU-0005375.
 PR 26-MAR-1996; 96AU-0008951.
 XX
 PA (AMRA-) AMRAD OPERATIONS PTY LTD.
 XX
 XX Koentgen F, Suess GM, Tarlington DM, Treutlein HR;
 PI WPI; 1997-489572/45.
 XX N-PSDB; AAT91589.
 DR
 XX New catalytic antibody precursors - comprising a B-cell surface
 PT molecule binding portion which can induce B-cell mitogenesis
 PT
 XX Example 11; Page 66-68; 109pp; English.
 PS
 XX This protein comprises growth factor TLHL, where L is the
 CC immunoglobulin binding entity from Peptostreptococcus magnus, H is
 CC residues 42-62 of hen egg lysozyme, and T represents the variable
 CC kappa light chain (see AW32483) from human myeloma protein LEN. It
 CC was expressed in E. coli DH10B cells utilising a DNA construct (see
 CC AAT91589) produced from LHL (see AAT91986), kappa (see AAT91590) and
 CC synthetic linker oligonucleotides, and was purified from total
 CC bacterial lysate via a strep-tag. The linker contains a cleavage
 CC site for tobacco etch virus (TEV) protease. TLHL was designed so
 CC that the kappa portion of the protein is cleaved by TEV protease so
 CC into 'T' and 'LHL' moieties. Production of catalytic antibodies to
 CC a specific antigen comprises administering to an animal a growth
 CC factor comprising an antigen capable of interacting with a B cell
 CC bound catalytic antibody. The antigen is fused to a B cell surface
 CC molecule binding protein for the antigen to be cleaved and for the
 CC remainder of the molecule to induce B cell mitogenesis (claimed).
 CC LHL crosslinks with surface immunoglobulin on B cells. This induces
 CC B cell activation and blast formation. Internalisation and
 CC processing of LHL leads to the presentation of the H peptide on
 CC MHC II. T cell recognition of MHC II with the H peptide signals the
 CC activated B cell to proliferate and undergo antibody class switching
 CC and secretion. The catalytic antibodies generated by the process
 CC can have diagnostic and therapeutic applications.
 XX
 SQ Sequence 342 AA;
 Query Match 39.5%; Score 618; DB 18; Length 342;
 Best Local Similarity 75.9%; Pred. No. 6e-39;
 Matches 126; Conservative 5; Mismatches 15; Indels 20; Gaps 1;
 QY 16 SEEVITKANLIFANGSTQTAEPFGTPEKATSEAYAYADTLKKDNGEYTVDVADKGYTLN 75
 Db 174 SAEVITKANLIFANGSTQTAEPFGTPEKATSEAYAYADTLKKDNGEYTVDVADKGYTLN 233
 QY 76 IKFAGKEKTPPEE-----PKEEVITKANLIYADGKTQTAEFGKTFE 115
 Db 234 IKFAGKEATNRNTDGDYDYGILQINSRWGLTSAEEVITKANLIFANGSTQTAEPFGKTFE 293
 QY 116 EATAEAYRYADALKKDNGETYVDVADKGYTLNKKFAGKEKTPPEPK 161
 Db 294 KATSEAYAYADTLKKDNGEYTVDVADKGYTLNKKFAGKESAWRHPQ 339
 RESULT 9
 ID AAY06909
 XX AAY06909 standard; Protein; 342 AA.
 AC AAY06909;
 XX
 DT 01-JUL-1999 (first entry)
 XX TLHL amino acid sequence.
 DE
 XX Growth factor precursor; B-cell surface; T cell surface; Cab; hepatitis;
 KW catalytic antibody; immunoglobulin; tumour necrosis factor; influenza;
 KW rheumatoid arthritis; toxic shock syndrome; viral docking receptor;
 KW human immune virus; tumour-specific antigen; amyloid plaque; myeloma;
 KW Alzheimer's disease; ICE; allergy; asthma; drug detoxification;
 KW autoimmune; inflammatory disease; gene therapy; protein L; P. magnus;
 XX hen egg lysozyme; HEL; LHL; TLHL.
 XX Peptostreptococcus magnus.
 OS Gallus sp.
 XX WO9915563-A1.
 PN 01-APR-1999.
 XX 18-SEP-1998; 98WO-AU00783.
 PF 19-SEP-1997; 97AU-0009306.
 PR (AMRA-) AMRAD OPERATIONS PTY LTD.
 XX Koentgen F, Suess GM, Tarlington DM, Treutlein HR;
 PI WPI; 1999-244394/20.
 XX N-PSDB; AAX34592.
 DR Growth factor precursor cleaved by antigen-specific catalytic
 XX antibody
 PT Example 11; Page 67-69; 101pp; English.
 PS The invention relates to a growth factor precursor that comprises B-cell
 XX surface binding part, T cell surface binding part, antigen cleavable
 CC by a catalytic antibody (CAB); and a peptide comprising heavy and light
 CC chains of immunoglobulin. When the antigen is cleaved the B cell surface
 CC part can interact with its target. The growth factor precursors are used
 CC to select B cells that produce Ag-specific CAB, and to generate CAB from
 CC such cells (by inducing mitogenesis, caused by the growth factor released
 CC by specific cleavage). The Ag-specific CAB can be directed against, e.g.
 CC tumour necrosis factor (treatment of rheumatoid arthritis or toxic shock
 CC syndrome); viral docking receptors (treatment of human immune virus,
 CC hepatitis and influenza infections); tumour-specific antigens; amyloid
 CC plaque (treatment of Alzheimer's disease or myeloma) or ICE (treatment of
 CC allergies such as asthma). CAB may also be used for drug detoxification,
 CC to treat autoimmune or inflammatory diseases and to eliminate
 CC environmental or industrial pollutants, such as plastics and petroleum.
 CC Particularly the growth factor precursors are produced by delivering the
 CC corresponding nucleic acid in a viral or other gene therapy vector. The
 CC present sequence represents the amino acid sequence of TLHL. The LHL is a
 CC growth factor comprising H flanked by two L molecules where L is a B cell
 CC surface molecule binding portion from protein L of P. magnus; H is a T
 CC cell surface molecule binding portion from hen egg lysozyme (HEL). TLHL
 CC is LHL linked to kappa chain via TEV (tobacco etch virus) sensitive
 CC peptide and g attached to N terminus region.
 XX
 SQ Sequence 342 AA;
 Query Match 39.5%; Score 618; DB 20; Length 342;
 Best Local Similarity 75.9%; Pred. No. 6e-39;
 Matches 126; Conservative 5; Mismatches 15; Indels 20; Gaps 1;
 QY 16 SEEVITKANLIFANGSTQTAEPFGTPEKATSEAYAYADTLKKDNGEYTVDVADKGYTLN 75
 Db 174 SAEVITKANLIFANGSTQTAEPFGTPEKATSEAYAYADTLKKDNGEYTVDVADKGYTLN 233
 QY 76 IKFAGKEKTPPEE-----PKEEVITKANLIYADGKTQTAEFGKTFE 115
 Db 234 IKFAGKEATNRNTDGDYDYGILQINSRWGLTSAEEVITKANLIFANGSTQTAEPFGKTFE 293
 QY 116 EATAEAYRYADALKKDNGETYVDVADKGYTLNKKFAGKEKTPPEPK 161
 Db 294 KATSEAYAYADTLKKDNGEYTVDVADKGYTLNKKFAGKESAWRHPQ 339

The antigen to be cleaved and for the remainder of the molecule to induce B cell mitogenesis (claimed). LHL.seq has the same biological activity as IHL, which crosslinks with surface immunoglobulin on B cells. This induces B cell activation and blast formation. Internalisation and processing of IHL leads to the presentation of the H peptide on MHC II. T cell recognition of MHC II with the H peptide signals the activated B cell to proliferate and undergo antibody class switching and secretion. The catalytic CC antibodies generated by the process can have diagnostic and therapeutic applications.

XX
SQ Sequence 198 AA;

Query Match 39.3%; Score 615; DB 18; Length 198;
Best Local Similarity 76.2%; Pred. No. 5.1e-39;
Matches 125; Conservative 5; Mismatches 14; Indels 20; Gaps 1

Qy 18 BEVTIKANLIIFANGSTQTAEFKGTPEKATSEAYAYADTLKKONGEYTVDVADKGTYLTNLIK 77
 ||||| |
Db 32 BEVTIKANLIIFANGSTQTAEFKGTPEKATSEAYAYADTLKKONGEYTVDVADKGTYLTNLIK 91
 ||||| |

Qy 78 FAGKEKTEE-----PKEEYTIKANLIYADGKTQTAEFKGTFEA 117
 ||||| |

Db 92 FAGKEATNRNTDGGTDYGILOINRWGLTSAEEYTIKANLIIFANGSTQTAEFKGTFEKA 151
 |:|:| |

Qy 118 TAEAYRYADALKKONGEYTVDVADKGTYLTNLIKFAEKETPEEPK 161
 |:|:| |

Db 152 TSEAYAYADTLKKONGEYTVDVADKGTYLTNLIKFAEKESAWRHPQ 195
 |:|:| |

RESULT 13
AA06910 ID AAY06910 standard; Protein; 198 AA.
XX
AC AAY06910;
XX
DT 01-JUL-1999 (first entry)
DE
XX
KW Amino acid sequence of a form of LHL growth factor (LHL.seq).
KW Growth factor precursor; B-cell surface; T cell surface; CAB; hepatitis;
KW catalytic antibody; immunoglobulin; tumour necrosis factor; influenza;
KW rheumatoid arthritis; toxic shock syndrome; viral docking receptor;
KW human immune virus; tumour-specific antigen; amyloid plaque; myeloma;
KW Alzheimer's disease; ICE; allergy; asthma; drug detoxification;
KW autoimmune; inflammatory disease; gene therapy; protein L; P. magnus;
KW hen egg lysozyme; HEL; LHL; LHL.seq.
XX
OS Peptostreptococcus magnus.
OS Gallus sp.
XX
PN W09915563-A1.
XX
PD 01-APR-1999.
XX
PF 18-SEP-1998; 98WO-AU00783.
XX
PR 19-SEP-1997; 97AU-0009306.
XX
PA (AMRA-) AMRAD OPERATIONS PTY LTD.
XX
PI Koentgen F, Suess GM, Tarlington DM, Treutlein HR;
XX
DR WFI; 1999-244394/20.
DR N-PSDB; AAX34593.
XX
PT Growth factor precursor cleaved by antigen-specific catalytic
XX antibody
XX
PS Example 3; Page 70-71; 101pp; English.
XX
CC The invention relates to a growth factor precursor that comprises B-cell
CC surface binding part, T cell surface binding part, antigen-cleavable

CC by a catalytic antibody (CAB); and a peptide comprising heavy and light
 CC chains of immunoglobulin. When the antigen is cleaved the B cell surface
 CC part can interact with its target. The growth factor precursors are used
 CC to select B cells that produce Ag-specific Cab, and to generate CAB from
 CC such cells (by inducing mitogenesis, caused by the growth factor released
 CC by specific cleavage). The Ag-specific CAB can be directed against, e.g.
 CC tumour necrosis factor (treatment of rheumatoid arthritis or toxic shock
 CC syndrome); viral docking receptors (treatment of human immune virus,
 CC hepatitis and influenza infections); tumour-specific antigens; amyloid
 CC plaque (treatment of Alzheimer's disease or myeloma) or IGF (treatment of
 CC allergies such as asthma). CAB may also be used for drug detoxification,
 CC to treat autoimmune or inflammatory diseases and to eliminate
 CC environmental or industrial pollutants, such as plastics and petroleum.
 CC Particularly the growth factor precursors are produced by delivering the
 CC corresponding nucleic acid in a viral or other gene therapy vector. The
 CC present sequence represents the amino acid sequence of the LHL growth
 CC factor carrying a N-terminal FLAG epitope and a C-terminal strep tag.
 CC L is a B cell surface molecule binding portion from protein L of
 CC P. magnus; H is a T cell surface molecule binding portion from hen egg
 CC lysozyme (HEL). LHL is a growth factor comprising H flanked by two L
 CC molecules.

XX Sequence 198 AA;

Query Match 39.3%; Score 615; DB 20; Length 198;
 Best Local Similarity 76.2%; Pred. No. 5.1e-39;
 Matches 125; Conservative 5; Mismatches 14; Indels 20; Gaps 1;

QY 18 EEVTIKANLIFANGSTQTAFFKGTSEAYAYADTLKKDNGEYTVADVADKGYTLNIK 77
 DB 32 EEVTIKANLIFANGSTQTAFFKGTSEAYAYADTLKKDNGEYTVADVADKGYTLNIK 91
 QY 78 FAKGKTEPTE-----PKEEVTIKANLIYADGKTOTAEFKGTFFEA 117
 DB 92 FAKGKATNRNTDSTGYILQINSRWGLTSAEEVTIKANLIFANGSTQTAFFKGTFEKA 151
 QY 118 TABAYAYADALKDNGEYTVADVADKGYTLNIKFAKGTPEEPK 161
 DB 152 TSEAYAYADTLKKDNGEYTVADVADKGYTLNIKFAKGESAWRHQ 195

RESULT 14
 AAW32480
 ID AAW32480 standard; Protein; 495 AA.

XX AAW32480;
 XX 27-MAR-1998 (first entry)
 DE Growth factor CATAB-TEV (catalytic antibody precursor).
 XX Catalytic antibody; growth factor; B-cell mitogenesis; LHL;
 KW L protein; hen egg lysozyme; kappa light chain.

XX Chimeric - Peptostreptococcus magnus.
 OS Chimeric - Gallus sp.
 OS Chimeric - Homo sapiens.

PN WO9735887-A1.

XX 02-OCT-1997.

XX 26-MAR-1997; 97WO-AU00194.

XX 27-FEB-1997; 97AU-0005375.

PR 26-MAR-1996; 96AU-0008951.

XX (AMRA-) AMRAD OPERATIONS PTY LTD.

XX Koentgen F, Suess GM, Tarlinton DM, Treutlein HR;

XX WPI; 1997-489572/45.

DR N-PSDB; AAT91587.

XX New catalytic antibody precursors - comprising a B-cell surface
 PT molecule binding portion which can induce B-cell mitogenesis
 XX Claim 12; Page 61-63; 109pp; English.

XX This polypeptide comprises CATAB-TEV, a synthetic growth factor
 CC composed of: (a) TLHL (see AAW91589), where L is the immunoglobulin
 CC binding entity of Peptostreptococcus magnus, H is residues 42-62 of
 CC hen egg lysozyme, and T represents the variable kappa light chain
 CC (see AAW32483) from human myeloma protein LEN; and (b) an additional,
 CC C-terminal kappa protein, the elements of CATAB-TEV being joined by
 CC linkers containing recognition sites for tobacco etch virus (TEV)
 CC protease. CATAB-TEV was designed so that the kappa portions of the
 CC protein are removed by TEV protease to release 'T' and 'LHL'
 CC moieties. Production of catalytic antibodies to a specific antigen
 CC comprises administering to an animal a growth factor comprising an
 CC antigen capable of interacting with a B cell bound catalytic
 CC antibody. The antigen is fused to a B cell surface molecule
 CC binding protein for the antigen to be cleaved and for the remainder
 CC of the molecule to induce B cell mitogenesis (claimed). LHL (see
 CC AAW32479) crosslinks with surface immunoglobulin on B cells. This
 CC induces B cell activation and blast formation. Internalisation
 CC and processing of LHL leads to the presentation of the H peptide on
 CC MHC II. T cell recognition of MHC II with the H peptide signals the
 CC activated B cell to proliferate and undergo antibody class switching
 CC and secretion. The catalytic antibodies generated by the process
 CC can have diagnostic and therapeutic applications.

XX Sequence 495 AA;

Query Match 39.2%; Score 614; DB 18; Length 495;
 Best Local Similarity 78.6%; Pred. No. 1.9e-38;
 Matches 125; Conservative 4; Mismatches 10; Indels 20; Gaps 1;

QY 16 SESEVTIKANLIFANGSTQTAFFKGTSEAYAYADTLKKDNGEYTVADVADKGYTLN 75
 DB 174 SAESEVTIKANLIFANGSTQTAFFKGTSEAYAYADTLKKDNGEYTVADVADKGYTLN 233
 QY 76 IKFAGKTEPTE-----PKEEVTIKANLIYADGKTOTAEFKGTFE 115
 DB 234 IKFAGKATNRNTDSTGYILQINSRWGLTSAEEVTIKANLIFANGSTQTAFFKGTFE 293
 QY 116 EATAEAYRYADALKDNGEYTVADVADKGYTLNIKFAKGE 154
 DB 294 KATSEAYAYADTLKKDNGEYTVADVADKGYTLNIKFAKGE 332

RESULT 15
 AAY06908
 ID AAY06908 standard; Protein; 495 AA.

XX AAY06908;

XX 01-JUL-1999 (first entry)

DE CATAB-TEV aminoacid sequence.

XX Growth factor precursor; B-cell surface; T cell surface; CAB; hepatitis;
 KW catalytic antibody; immunoglobulin; tumour necrosis factor; influenza;
 KW rheumatoid arthritis; toxic shock syndrome; viral docking receptor;
 KW human immune virus; tumour-specific antigen; amyloid plaque; myeloma;
 KW Alzheimer's disease; IGE; allergy; asthma; drug detoxification;
 KW autoimmune; inflammatory disease; gene therapy; protein L; P. magnus;
 KW hen egg lysozyme; HEL; LHL; CATAB-TEV; tobacco etch virus; TLHL.

OS Peptostreptococcus magnus.

OS Gallus sp.

XX WO9915563-A1.

XX 01-APR-1999.

PD

XX

GenCore version 5.1.3
Copyright (c) 1993 - 2002 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: October 29, 2002, 09:25:26 ; Search time 10.571 Seconds
(without alignments)
1589.653 Million cell updates/sec

Title: US-08-325-278B-3

Perfect score: 2235

Sequence: 1 AVENKETPTPTDSEEV.....GVDGVWYDDATKTFVTVM 434

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 105224 seqs, 38719550 residues

Total number of hits satisfying chosen parameters: 105224

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt_40.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	708	81.7	593	1 SPG2_STRSP	P19009 streptococ
2	683	30.6	448	1 SPG1_STRSP	P06654 streptococ
3	188.5	8.4	1183	1 CNA_STAAU	Q53654 staphylococ
4	156	7.0	915	1 NFM_HUMAN	P07197 homo sapien
5	155	6.9	555	1 EDCB_DAUCA	P20075 daucus caro
6	151.5	6.8	1637	1 MRSP_STAAU	P80544 staphylococ
7	149	6.7	1433	1 SUBE_BACSU	P16397 bacillus su
8	148.5	6.6	2647	1 ABP2_HUMAN	P21333 homo sapien
9	148	6.6	1233	1 YFI6_YEAST	P43597 saccharomyc
10	146.5	6.6	1337	1 DEXT_STRDO	P39653 streptococ
11	144	6.4	2334	1 WAPA_BACSU	Q07833 bacillus su
12	143.5	6.4	848	1 NFM_MOUSE	P08553 mus musculu
13	142.5	6.4	814	1 SLAL_BACAN	P49051 bacillus an
14	141.5	6.3	845	1 NFM_RAT	P12839 rattus norv
15	141	6.3	2459	1 MAPB_MOUSE	P15205 rattus norv
16	138.5	6.2	2464	1 MAPB_MOUSE	P14873 mus musculu
17	138	6.2	2468	1 MAPB_HUMAN	P46821 homo sapien
18	136.5	6.1	1500	1 SPSP_STRGN	P16952 streptococ
19	134.5	6.0	1940	1 MYH3_CHICK	P02565 gallus gall
20	134	6.0	1935	1 MYSS_CYPCA	Q90339 cyprinus ca
21	133.5	6.0	810	1 NFM_BOVIN	Q07788 bos taurus
22	133	6.0	505	1 FLJB_SALTY	P52616 salmonella
23	132.5	5.9	599	1 TBB2_NEIMB	Q06988 netisseria m
24	131.5	5.9	857	1 NFM_CHICK	P16053 gallus gall
25	131.5	5.9	1018	1 FNSA_STAAU	P47338 staphylococ
26	131.5	5.9	2116	1 MYB2_DICDI	P08799 dictyosteli
27	131	5.9	6669	1 NEBU_HUMAN	P20929 homo sapien
28	130	5.8	1277	1 IF3X_YEAST	Q03690 saccharomyc
29	129.5	5.8	1104	1 COLA_CLOPE	P43153 clostridium
30	129.5	5.8	1938	1 MYHA_RABIT	Q28441 oryctolagus
31	129	5.8	1176	1 SLAP_BACSH	P38537 bacillus sp
32	127	5.7	929	1 CAIC_NOTVI	Q91145 notophthalm
33	125.5	5.6	827	1 CSG_HALVO	P25062 halobacteri

ALIGNMENTS

RESULT 1

SPG2_STRSP STANDARD; PRT; 593 AA.
AC P19009;
DT 01-FEB-1991 (Rel. 17, Created)
DT 01-FEB-1991 (Rel. 17, Last sequence update)
DT 15-JUL-1999 (Rel. 38, Last annotation update)
DE Immunoglobulin G binding protein G precursor (IGS binding protein G).
GN SPC.
OS Streptococcus sp. (strain G148).
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Streptococcaceae;
OC Streptococcus.
OX NCBI_TaxID=1306;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=G148;
RX MEDLINE=88029445; PubMed=3665928;
RA Olsson A., Eliasson M., Guss B., Nilsson B., Hellman U.,
RA Lindberg M., Uhlen M.;
RT "Structure and evolution of the repetitive gene encoding
RT streptococcal protein G.";
RL Eur. J. Biochem. 168:319-324(1987).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=GX7805;
RX MEDLINE=88015586; PubMed=3658689;
RA Filpula D., Alexander P., Fahnestock S.R.;
RT "Nucleotide sequence of the protein G gene from Streptococcus GX7805,
RT and comparison to previously reported sequences.";
RL Nucleic Acids Res. 15:7210-7210(1987).
RN [3]
RP SEQUENCE OF 114-593 FROM N.A.
RC STRAIN=G148;
RX MEDLINE=86300657; PubMed=3017704;
RA Guss B., Eliasson M., Olsson A., Uhlen M., Frej A.-K., Joernvall H.,
RA Flock J.-I., Lindberg M.;
RT "Structure of the IgG-binding regions of streptococcal protein G.";
RL EMBO J. 5:1567-1575(1986).
RN [4]
RP STRUCTURE BY NMR OF 371-427.
RC STRAIN=G148;
RX MEDLINE=98290449; PubMed=9628485;
RA Malakauskas S.M., Mayo S.L.;
RT "Design, structure and stability of a hyperthermophilic protein
RT variant.";
RL Nat. Struct. Biol. 5:470-475(1998).
CC -!- SUBCELLULAR LOCATION: Type I membrane protein. Cell wall.
CC -!- SIMILARITY: TO OTHER STREPTOCOCCAL AND STAPHYLOCOCCAL PROTEINS
CC IN THE REGION OF THE MEMBRANE ANCHOR.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>)

34 125.5 5.6 1053 1 SLPM_BACBR
35 125 5.6 644 1 NFM_RABIT
36 124.5 5.6 3135 1 S230_PLAFO
37 124 5.5 705 1 CWBA_BACSU
38 124 5.5 706 1 NUCL_MOUSE
39 124 5.5 852 1 CSG_HALHA
40 124 5.5 1116 1 SLPH_BACBR
41 124 5.5 1902 1 P1P_LACLC
42 124 5.5 1902 1 P3P_LACLC
43 124 5.5 3063 1 CAIC_HUMAN
44 123.5 5.5 589 1 HIBN_XENLA
45 123.5 5.5 934 1 EAE_ECO57

P06546 bacillus br
P54938 oryctolagus
Q08372 plasmodium
Q02113 bacillus su
P09405 mus musculu
P08198 halobacteri
P38538 bacillus br
P16371 lactococcus
P15292 lactococcus
Q99715 homo sapien
P06180 xenopus lae
P43261 escherichia

or send an email to license@isb-sib.ch.

```

CC  EMBL; X06173; CAA29540.1; -
CC  EMBL; Y00428; CAA68489.1; -
DR  EMBL; X04015; CAA27638.1; -
DR  PIR; S00128; S00128.
DR  PIR; A26314; A26314.
DR  PDB; 1GB4; 22-JUL-98.
DR  InterPro; IPR002988; GA.
DR  InterPro; IPR001899; Gram_pos_anchor.
DR  InterPro; IPR000724; IgG_bind_B.
DR  Pfam; PF01468; GA; 3.
DR  Pfam; PF00746; Gram_pos_anchor; 1.
DR  Pfam; PF01378; IgG_binding_B; 3.
DR  PRINTS; PR00015; GPOSANCHOR.
DR  PROSITE; PS00343; GRAM_POS_ANCHORING; 1.
KW  IgG-binding protein; Repeat; Transmembrane; Cell wall; Signal;
KW  3D-structure.
FT  SIGNAL 1 33
FT  CHAIN 34 593 IMMUNOGLOBULIN G BINDING PROTEIN G.
FT  DOMAIN 34 567 EXTRACELLULAR (POTENTIAL).
FT  TRANSMEM 568 588 MEMBRANE ANCHOR.
FT  DOMAIN 589 593 CYTOPLASMIC (POTENTIAL).
FT  DOMAIN 69 111 ALA-RICH.
FT  DOMAIN 104 290 3 X 37 AA REPEATS.
FT  REPEAT 104 140 1-1.
FT  REPEAT 179 215 1-2.
FT  REPEAT 254 290 1-3.
FT  DOMAIN 303 427 2 X 55 AA REPEATS.
FT  REPEAT 303 357 2-1.
FT  REPEAT 373 427 2-2.
FT  DOMAIN 531 555 5 X 5 AA REPEATS OF [DE]-D-A-K-K.
FT  DOMAIN 559 564 CONSERVED IN GRAM-POSITIVE COCCI SURFACE
FT  PROTEINS.
SQ  SEQUENCE 593 AA; 63291 MW; 048BAA760D5B2920 CRC64;
Query Match 31.7%; Score 708; DB 1; Length 593;
Best Local Similarity 48.3%; Pred. No. 1.2e-31;
Matches 205; Conservative 28; Mismatches 131; Indels 60; Gaps 17;
QY 45 ATSEAYAYADTLKDNCEYIVDVADKGYTLNIFAGKEKTPPEKPEVITKANLIYA-DG 103
DQ 101 ADALAKAKADALKEFN-KYGVSDYYKLNINNAKTVEGSKDLQAOVVEAKKARISEATDG 159
QY 104 -----KTQTAEFKGTFFEEATAEAYRYADALKDNGE---YTVDVADKGYTLNIFAGKEK 155
DQ 160 LSDFLKSQTP-----AEDTVKSIEIAEAKVLANRELDKYGVSDYHKNLINNAKTVEGVK 213
QY 156 TPEEPKEEVIKANLIYA-DG-----KTQTAEPKGTFFEEATAEAYRYAD---LLAKENGK 206
DQ 214 DLQAOVVEAKKARISEATDGLSDFLKSQTP-----AEDTVKSIEIAEAKVLANRELDK 267
QY 207 YTVDVADKGYTLNIFAGKEKTPPE-----PKEEVIKANLIYADKGTQTAEFKGTFAE 260
DQ 268 YGVSDYYKLNINNAKTVEGKALIDETLAALPKTD-TYK---LILNGKTLKGE---TTTE 320
QY 261 A--TAEAYRYADLILAKNGKGYTADLEDGGYTIINRFAGKKVDKPE-----EPMQTY 310
DQ 321 AVDAATAEKVKQYANDNG-----VDGEWYDDATKFTFTVTEKPEVIDASELTPAVTY 374
QY 311 KLILNGKTLKGETTTEAVDAATAEKVKQYANDNGVDGEWYDDATKFTFTVTEKPEVIDA 370
DQ 375 KLVIKNGKTLKGETTTEAVDAATAEKVKQYANDNGVDGEWYDDATKFTFTVTEKPEVIDA 434
QY 371 SELTPAVTYTKLVINGKTLKGETTTTAKVDAETAETAKFAKQYANDNGVDGVTYDDATKFT 430
DQ 435 SELTPAVTYTKLVINGKTLKGETTTTAKVDAETAETAKFAKQYANDNGVDGVTYDDATKFT 494
QY 431 VTEM 434
DQ 495 VTEM 498

```

```

RESULT 2
SPGI_STRSP
ID SPGI_STRSP STANDARD; PRT; 448 AA.
AC P06654;
DT 01-JAN-1988 (Rel. 06, Created)
DT 01-JAN-1988 (Rel. 06, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Immunoglobulin G binding protein G precursor (IGG binding protein G).
SPG.
OS Streptococcus sp. (Lancefield group G).
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Streptococcaceae;
OC Streptococcus.
OX NCBI_TaxID=1306;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=86304178; PubMed=3745123;
RA Fannescock S.R., Alexander P., Nagle J., Filpula D.;
RT "Gene for an immunoglobulin-binding protein from a group G
RT streptococcus.";
RL J. Bacteriol. 167:870-880(1986).
RN [2]
RP X-RAY CRYSTALLOGRAPHY (2.07 ANGSTROMS) OF 228-282.
RX MEDLINE=94213848; PubMed=8161530;
RA Gallagher T., Alexander P., Bryan P., Gilliland G.L.;
RT "Two crystal structures of the B1 immunoglobulin-binding domain of
RT streptococcal protein G and comparison with NMR.";
RL Biochemistry 33:4721-4729(1994).
RN [3]
RP X-RAY CRYSTALLOGRAPHY (1.1 ANGSTROMS) OF 293-351.
RX MEDLINE=95055731; PubMed=7966308;
RA Derrick J.P., Wigley D.B.;
RT "The third IgG-binding domain from streptococcal protein G. An
RT analysis by X-ray crystallography of the structure alone and in a
RT complex with Fab.";
RL J. Mol. Biol. 243:906-918(1994).
RN [4]
RP X-RAY CRYSTALLOGRAPHY (3.5 ANGSTROMS) OF 297-352.
RX MEDLINE=95308043; PubMed=7788293;
RA Sauer-Briksson A.E., Kleywegt G.J., Uhlen M., Jones T.A.;
RT "Crystal structure of the C2 fragment of streptococcal protein G in
RT complex with the FC domain of human IgG.";
RL Structure 3:265-278(1995).
RN [5]
RP X-RAY CRYSTALLOGRAPHY (1.1 ANGSTROMS) OF 292-352.
RA Butterworth S., Lamin V.S., Wigley D.B., Derrick J.P., Wilson K.S.;
RT Submitted (APR-1997) to the PDB data bank.
RN [6]
RP STRUCTURE BY NMR OF 298-351.
RX MEDLINE=91335209; PubMed=1871600;
RA Gronenborn A.M., Filpula D.R., Essig N.Z., Achari A., Whitlow M.,
RA Wingfield P.T., Clore G.M.;
RT "A novel, highly stable fold of the immunoglobulin binding domain of
RT streptococcal protein G.";
RL Science 253:657-661(1991).
CC -|- FUNCTION: BINDS TO THE CONSTANT FC REGION OF IGG WITH HIGH
CC AFFINITY.
CC -|- SUBCELLULAR LOCATION: Type I membrane protein. Cell wall.
CC -|- SIMILARITY: TO OTHER STREPTOCOCCAL AND STAPHYLOCOCCAL PROTEINS
CC IN THE REGION OF THE MEMBRANE ANCHOR.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: M13825; AAA03664.1; -
CC PIR: A24496; A24496.
CC PDB: 1PGA; 30-APR-94.
CC PDB: 1PGB; 30-APR-94.
CC PDB: 1PGX; 15-JUL-92.

```

```
DR PDB; 1IGC; 03-JUN-95.
DR PDB; 1IGD; 01-NOV-94.
DR PDB; 2IGD; 29-JUL-98.
DR PDB; 1FCC; 20-JUL-95.
DR PDB; 1GB1; 15-APR-93.
DR InterPro: IPR002988; GA.
DR InterPro: IPR001899; Gram_pos_anchor.
DR InterPro: IPR000724; Igg_bind_B.
DR Pfam: PF00746; Gram_pos_anchor; 1.
DR Pfam: PF01378; Igg_binding_B; 2.
DR PRINTS; PR00015; GP0SANCHOR.
DR PROSITE; PS00343; GRAM_POS_ANCHORING; 1.
KW Igg-binding protein; Repeat; Transmembrane; Cell wall; Signal;
KW 3D-structure.
FT SIGNAL 1 33
FT CHAIN 34 448 IMMUNOGLOBULIN G BINDING PROTEIN G.
FT DOMAIN 34 422 EXTRACELLULAR.
FT TRANSMEM 423 443
FT DOMAIN 444 448
FT CYTOPLASMIC.
FT DOMAIN 69 111 ALA-RICH.
FT DOMAIN 104 215 2 X 37 AA REPEATS.
FT REPEAT 104 140 1-1.
FT REPEAT 179 215 1-2.
FT DOMAIN 228 352 2 X 55 AA REPEATS.
FT REPEAT 228 282 2-1.
FT REPEAT 298 352 2-2.
FT DOMAIN 386 410 5 X 5 AA REPEATS OF [DE]-D-A-K-K.
FT DOMAIN 414 419 CONSERVED IN GRAM-POSITIVE COCCI SURFACE
FT PROTEINS.
SQ SEQUENCE 448 AA; 47567 MW; A0759060C8F9E6CA CRC64;

Query Match 30.6%; Score 683; DB 1; Length 448;
Best Local Similarity 53.2%; Pred. No. 1.9e-30;
Matches 176; Conservative 15; Mismatches 92; Indels 48; Gaps 9;

QY 130 KDMGEYTVADVADKGYTLNIFAGKEKTPPEPKKEVITKANLIYAD-GKTTAEFGKTFEE 188
Db 45 RRGSELT-----NLGNSETTLALRNEESATDLTAAAVADVAAAAENAGA 92
QY 189 ATAERYAYADLLA-----KENGKYTVADVADKGYTLNIFAGKEKTPPEPKKEVITKAN 241
Db 93 AAWEAAADALAKAKADALKEFNKYGVSQYKLNINNAKTVEGIDKDAQVVESSAKKAR 152
QY 242 LIYA-DG-----KTOTAEFGKTAFAEAYRVAD-----LLAKENGKYTADLEDGGYITNI 292
Db 153 ISEATDGLSDFLSKQTP-----AEDTVKSIELAEAKVLANRELDKYGV-----SDYHKNL 202
QY 293 RFAGKKYVDEKPE-----EPMDYKYLINGKTLKGETTTEAVDAATAEKVFKQYAND 343
Db 203 INNAKTVEGVKELIDEILAAALPKDVTYKYLINGKTLKGETTTEAVDAATAEKVFKQYAND 262
QY 344 NGVDGEYTVDDATKFTFTVEKPEVIDASELTTPAVTTVKLVINGKTLKGETTTTAVDAETA 403
Db 263 NGVDGEYTVDDATKFTFTVEKPEVIDASELTTPAVTTVKLVINGKTLKGETTTTAVDAETA 322
QY 404 EKAFQKYANDNGVDGVWTVDDATKFTFTVTEM 434
Db 323 EKAFQKYANDNGVDGVWTVDDATKFTFTVTEM 353

RESULT 3
CNA_STAAU
ID CNA_STAAU STANDARD; PRT; 1183 AA.
AC Q53634;
DT 15-DEC-1998 (Rel. 37, Created)
DT 15-DEC-1998 (Rel. 37, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Collagen adhesin precursor.
GN CNA.
OS Staphylococcus aureus.
OC Bacteria; Firmicutes; Bacillus/Clostridium group;
OC Bacillus/Staphylococcus group; Staphylococcus.

NCBI_TaxID=1280;
[1]
SEQUENCE FROM N.A.
RX STRAIN=FDA 574;
RX MEDLINE=92165839; PubMed=1311320;
RA Patti J.M., Jonsson H., Guss B., Switalski L.M., Wiberg K.,
RA Lindberg M., Hoeoek M.;
RT "Molecular characterization and expression of a gene encoding a
RT Staphylococcus aureus collagen adhesin.";
RL J. Biol. Chem. 267:4766-4772(1992).
RN [2]
ERRATUM.
RA Patti J.M., Jonsson H., Guss B., Switalski L.M., Wiberg K.,
RA Lindberg M., Hoeoek M.;
RL J. Biol. Chem. 269:11672-11672(1994).
RN [3]
COLLAGEN-BINDING DOMAIN.
RX STRAIN=FDA 574;
RX MEDLINE=94032261; PubMed=8218209;
RA Patti J.M., Boles J.O., Hoeoek M.;
RT "Identification and biochemical characterization of the ligand
RT binding domain of the collagen adhesin from Staphylococcus aureus.";
RL Biochemistry 32:11428-11435(1993).
RN [4]
X-RAY CRYSTALLOGRAPHY (2.0 ANGSTROMS) OF 169-318.
RX MEDLINE=97475225; PubMed=9334749;
RA Symersky J., Patti J.M., Carson M., House-Pompeo K., Teale M.,
RA Moore D., Jin L., Schneider A., Delucas L.J., Hoeoek M.,
RA Narayana S.V.L.;
RT "Structure of the collagen-binding domain from a Staphylococcus
RT aureus adhesin.";
RL Nat. Struct. Biol. 4:833-838(1997).
CC -!- FUNCTION: MEDIATES ATTACHMENT OF STAPHYLOCOCCAL CELLS TO
CC COLLAGEN-CONTAINING SUBSTRATA.
CC -!- SUBCELLULAR LOCATION: Type I membrane protein. Cell wall.
CC -!- SIMILARITY: TO OTHER STREPTOCOCCAL AND STAPHYLOCOCCAL PROTEINS
CC IN THE REGION OF THE MEMBRANE ANCHOR.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch)
CC -----
EMBL; M81736; AAA20874.1; -
DR PDB; 1AMX; 24-JUN-98.
DR InterPro: IPR001899; Gram_pos_anchor.
DR PROSITE; PS00343; GRAM_POS_ANCHORING; FALSE_NEG.
KW Signal; Repeat; Transmembrane; Cell wall; 3D-structure.
FT SIGNAL 1 29 POTENTIAL.
FT CHAIN 30 1183
FT DOMAIN 30 1157 EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 1158 1177 MEMBRANE ANCHOR (POTENTIAL).
FT DOMAIN 1178 1183 CYTOPLASMIC (POTENTIAL).
FT DOMAIN 151 318 COLLAGEN-BINDING.
FT DOMAIN 533 1093 3 X 187 AA APPROXIMATE TANDEM REPEATS.
FT DOMAIN 1093 1157 LYS/PRO-RICH (CELL WALL-SPANNING).
FT DOMAIN 1151 1156 CONSERVED IN GRAM-POSITIVE COCCI SURFACE
FT PROTEINS.
FT REPEAT 533 719 B1.
FT REPEAT 720 906 B2.
FT REPEAT 907 1093 B3.
SQ SEQUENCE 1183 AA; 133066 MW; B6A1CC072E575D76 CRC64;

Query Match 8.4%; Score 188.5; DB 1; Length 1183;
Best Local Similarity 23.8%; Pred. No. 0.0032;
Matches 136; Conservative 61; Mismatches 166; Indels 209; Gaps 38;

QY 2 VENKEETPETPETDSE-----EVTYKANLIFANG-STQTAEFKGTTEKAT 46
Db 525 VTNTEPIETTSISGKVDKQDQKRPKYSVN---LLANGKVKYTLT-----VT 574
```



```

DR InterPro; IPR001899; Gram_pos_anchor.
DR Pfam; PF00746; Gram_pos_anchor; 1.
DR - PROSITE; PS00343; GRAM_POS_ANCHORING; FALSE_NEG.
KW Cell wall; Transmembrane; Antibiotic resistance; Glycoprotein; Repeat;
KW Signal.
FT SIGNAL 1 48 POTENTIAL.
FT CHAIN 49 1637 METHICILLIN-RESISTANT SURFACE PROTEIN.
FT DOMAIN 1301 1582 141 X 2 AA TANDEM REPEATS OF D-[SAG].
FT DOMAIN 1598 1603 CONSERVED IN GRAM-POSITIVE COCCI SURFACE
FT PROTEINS.
SQ SEQUENCE 1637 AA; 174573 MW; 75BE9ADB469BD309 CRC64;

Query Match 6.8%; Score 151.5; DB 1; Length 1637;
Best Local Similarity 24.4%; Pred. No. 0.46;
Matches 111; Conservative 34; Mismatches 200; Indels 109; Gaps 17;

QY 3 ENKEETPETTETSEEEVITKANLIIFANGSTQTAEPKFGFEKATSAAYAYADTLKKDNGE 62
DQ 87 EQVDVTKDITTEQASTEE---KAN-----TTEQASTEERKADTTTEQATTEAPKAEG- 133
QY 63 YTVADVADKGYTLNKPAGKEKTPPEEKVEVITKANLIYADGKTQTAEFKGTREATAEAY 122
DQ 134 -----TDKVTETEAAPKAETDKAT-----TEEPKAEETDKATEEAP 170
QY 123 RYADALKKDNGETYVDVADKGYTLNKFKA---GKEKTPPEKPEEVITKANLIYA-----D 174
DQ 171 K-----TEETDKATTEEAPAAEETSKAATEEAPKAETSKAATEEAPKAEEET 217
QY 175 GKQTAEFKGTTEE---ATPEAYRYADLLAKENGKTYVDVADKGYTLNKFAGKEKTPPEE 231
DQ 218 EKTATEEAPKTEETDKVETEAPK-----AEETSKATEKAPKAETN-----KVETEAP 267
QY 232 PKEEVTIKANLIYADGKTQTAEPKFGFAEATA-----EAYRYADLLAKENGKTYADLDG 286
DQ 268 PAEEETNKA-----ATETPAVEDTNKASNSNAOPSETQTVVDTVAKDLKKSEVTE-- 321
QY 287 GYTINIRFAGKKVDEKPEEPMPTDKYLILNGKTLKGTTTTEAVDAAYAEKVKQYANDNGV 346
DQ 322 -----AEKAEIKVLKPKDISNL-----SNEEIKKIALSEVLKETANKENA 361
QY 347 DGEWYTDATKTVTEKPEVIDASELTPAVTYTKLVINGKTLKGTTTKKAVDAETAEK- 405
DQ 362 OPRATFRVSSNARTNNVNSATLRAAODTVTKGTGNTGAHDIIHKTYKEEFPNEG 421
QY 406 ---AFKQYANDN-GVDGVMWYDDA---TKTFTVT 432
DQ 422 TLTAFTNPNPNNGTKGALEYNDKIDFNKDFTTIT 455

RESULT 7
SUBF_BACSU STANDARD; PRT; 1433 AA.
AC P16397;
DT 01-AUG-1990 (Rel. 15, Created)
DT 01-FEB-1996 (Rel. 33, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Bacillopeptidase F precursor (EC 3.4.21.-) (Esterase) (RP-I protease)
DE (90 kDa serine proteinase).
GN BPR OR BPF.
OS Bacillus subtilis.
OC Bacteria; Firmicutes; Bacillus/Clostridium group;
OC Bacillus/Staphylococcus group; Bacillus.
OX NCBI_TaxID=1423;
RN [1]
RP SEQUENCE FROM N.A., AND SEQUENCE OF 195-222.
RX MEDLINE=90170864; PubMed=2106512;
RA Sloma A., Rufo G.A. Jr., Rudolph C.F., Sullivan B.J., Theriault K.A.,
RA Pero J.;
RT "Bacillopeptidase F of Bacillus subtilis: purification of the protein
RT and cloning of the gene.";
RL J. Bacteriol. 172:1470-1477(1990).
RN [2]
RP REVISIONS.

```

```

RC STRAIN=168;
RX MEDLINE=90368623; PubMed=2118514;
RA Sloma A., Rufo G.A. Jr., Rudolph C.F., Sullivan B.J., Theriault K.A.,
RA Pero J.;
RL J. Bacteriol. 172:5520-5521(1990).
RN [3]
RP SEQUENCE FROM N.A.
RX MEDLINE=90216713; PubMed=2108961;
RA Wu X.-C., Nathoo S., Pang A.S.-H., Carne T., Wang S.-L.;
RT "Cloning, genetic organization, and characterization of a structural
RT gene encoding bacillopeptidase F from Bacillus subtilis.";
RL J. Biol. Chem. 265:6845-6850(1990).
RN [4]
RP SEQUENCE OF 1-211 FROM N.A.
RX MEDLINE=89008108; PubMed=3139638;
RA Beall B., Lowe M., Lutkenhaus J.;
RT "Cloning and characterization of Bacillus subtilis homologs of
RT Escherichia coli cell division genesftsZ andftsA.";
RL J. Bacteriol. 170:4855-4864(1988).
RN [5]
RP SEQUENCE OF 1410-1433 FROM N.A.
RC STRAIN=168 / MARBURG;
RX MEDLINE=90174995; PubMed=2106671;
RA Masuda E.S., Anaguchi H., Sato T., Takeuchi M., Kobayashi Y.;
RT "Nucleotide sequence of the sporulation gene spoIIIGA from Bacillus
RT subtilis.";
RL Nucleic Acids Res. 18:657-657(1990).
RN [6]
RP SEQUENCE OF 195-219.
RC STRAIN=NAITTO 16;
RA Kato T., Yamagata Y., Arai T., Ichishima E.;
RT "Purification of a new extracellular 90-kDa serine proteinase with
RT isoelectric point of 3.9 from Bacillus subtilis (natto) and
RT elucidation of its distinct mode of action.";
RL Biosci. Biotechnol. Biochem. 56:1166-1168(1992).
CC -!- SUBCELLULAR LOCATION: Secreted.
CC -!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S8; ALSO KNOWN AS THE
CC SUBTILASE FAMILY.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; M29035; AAA62679.1; -
CC EMBL; J05400; AAA83362.1; -
CC EMBL; M22630; AAA22458.1; -
CC EMBL; X17344; CAA35224.1; -
CC EMBL; Z99111; CAB13403.1; -
CC EMBL; Z99112; CAB13404.1; -
CC PIR; A35131; A35131.
CC PIR; A35750; A35750.
CC PIR; A36734; A36734.
CC HSP; P00782; 2SPT.
CC MEROPS; S08.017; -.
CC Subtilist; BG10233; bpr.
CC InterPro; IPR000209; Peptidase_S8.
CC Pfam; PF00082; Peptidase_S8; 1.
CC PRINTS; PR00723; SUBTILISIN.
CC PROSITE; PS00136; SUBTILASE_ASP; FALSE_NEG.
CC PROSITE; PS00137; SUBTILASE_HIS; 1.
CC PROSITE; PS00138; SUBTILASE_SER; 1.
KW Hydrolase; Serine protease; Zymogen; Signal; Complete proteome.
FT SIGNAL 1 30
FT PROPEP 31 194 POTENTIAL.
FT CHAIN 195 755 BACILLOPEPTIDASE F.
FT PROPEP 756 1433 POTENTIAL.
FT ACT_SITE 227 227 CHARGE RELAY SYSTEM (BY SIMILARITY).
FT ACT_SITE 274 274 CHARGE RELAY SYSTEM (BY SIMILARITY).
FT ACT_SITE 452 452 CHARGE RELAY SYSTEM (BY SIMILARITY).

```

```
FT CONFLICT 219 219 T -> A (IN REF. 6).
FT CONFLICT 393 393 A -> V (IN REF. 3).
FT CONFLICT 829 834 KHONKA -> N (IN REF. 3).
FT CONFLICT 836 841 QQVLP -> RRLYS (IN REF. 3).
FT CONFLICT 844 852 AQSVVETG -> FCRSRKSV (IN REF. 3).
FT CONFLICT 853 1433 MISSING (IN REF. 3).
SQ SEQUENCE 1433 AA: 154577 MW: 98DF6846897807C9 CRC64;

Query Match 6.7%; Score 149; DB 1; Length 1433;
Best Local Similarity 21.6%; Pred. No. 0.54;
Matches 116; Conservative 59; Mismatches 187; Indels 176; Gaps 27;

QY 2 VENKEETPETPDSSEEVIRKANLIFANGSTQTAF - KGTFEKATSEAYAYADTLKKDN 60
Db 815 VNPKKAPSAANTAVKHONKAIQOVLPKKAQSVVETGKSTYSDQSTGOY----TLKHA 870
QY 61 GEYTDVADKGYTLNFKAGKEKTPPE-EPKEEYVIRKANLYADGKQTAFKFTFEAT- 118
Db 871 GDYTLMAEAYGY-----QSTQKQVSLKTDQTQANFTLEMKKGT--LKGTIVINKTT 920
QY 119 -----AEARYAD-----ALKKNGEYTVDVADKGYTLNFKAG--KEKTPPEKPEEVT 165
Db 921 GEPVTSVYVVEDAAVEPAMTWKGEYMLEAYEGATIKVAAPGYSDSEFVELKGDVT 980
QY 166 -----IKANLYADGKTQTA----- 180
Db 981 KETALPKFVGYPGEIAYDGTAEANANSYFAAGNGWAKMTLADGDKGMLTGLFRFWDT 1040
QY 181 -----EPKGTFEAT-----AEARYADLLAKENGKYT-VDVADKG- 215
Db 1041 EFPDPGTEPKFVYDGTGDKGAPGKGIAPFNAELR-----NGETWKVDSLSSGI 1092
QY 216 -----YTLNFKAGKEKTPPEKPEEVIKA--NLIVADGKTO-----TAEF 254
Db 1093 MVDKDFLYVIQSKPDYSPGLAMDETGQNSGRNQWIDGKQPGKADGNVIMRALVDY 1152
QY 255 KGTFAEATA---BARYAD--LLAKENGKYTADLEDGGYTNIRFAGKVDKPEEPMDTY 310
Db 1153 EAAVPEITSPDKSYNTKDSVTYKQNA-----SPGTTVHIYNGEKEAGETKAAADGTF 1205
QY 311 K--LILNGTLKGETTTEAVDAATAEKVQYANDNGVGEWYDDAKYTFVT---EKP 365
Db 1206 HAGIILN---KGENEL---TATA-----STDNGT-----TDASSPITVTLDOEKP 1244
QY 366 EVI-----DASELTPAVTVTKLVINGKTLK-----GETTKAVDAETAERAKFOYAND 413
Db 1245 ELTLDNPKDGGKYNKETLTVKGAVSDNLDKVKVNGKKATVADGYSARILLENGRNE 1302

RESULT 8
ABP2_HUMAN STANDARD; PRT; 2647 AA.
AC P21333;
DT 01-MAY-1991 (Rel. 18, Created)
DT 01-MAY-1991 (Rel. 18, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Endothelial actin-binding protein (ABP-280) (Nonmuscle filamin)
DE (Filamin 1) (Alpha-filamin).
GN FLN1 OR FLN.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.
RX MEDLINE=90361737; PubMed=2391361;
RA Gorlin J.B., Yamin R., Egan S., Stewart M., Stossel T.P.,
RA Kwiatkowski D.J., Hartwig J.H.;
RT "human endothelial actin-binding protein (ABP-280, nonmuscle
RL filamin): a molecular leaf spring.";
RN [2]
RP SEQUENCE FROM N.A.
```

```
RX MEDLINE=96311563; PubMed=8733135;
RA Chen E.Y., Zollio M., Mazzarella R.A., Ciccocioppa A., Chen C.-N.,
RA Zuo L., Heiner C., Burrough F.W., Ripetto M., Schlessinger D.,
RA D'Urso M.;
RT "Long-range sequence analysis in Xq28: thirteen known and six
RT candidate genes in 219.4 kb of high GC DNA between the RCP/GCP and
RT G6PD loci.";
RL Hum. Mol. Genet. 5:659-668(1996).
RN [3]
RP SEQUENCE OF 1658-1772 FROM N.A.
RX MEDLINE=93357748; PubMed=7689010;
RA Maestrini E., Patrosso C., Mancini M., Rivella S., Rocchi M.,
RA Repetto M., Villa A., Frattini A., Zoppe M., Vezzoni P.,
RA Toniolo D.;
RT "Mapping of two genes encoding isoforms of the actin binding protein
RT ABP-280, a dystrophin like protein, to Xq28 and to chromosome 7.";
RL Hum. Mol. Genet. 2:761-766(1993).
CC -!- FUNCTION: PROMOTES ORTHOGONAL BRANCHING OF ACTIN FILAMENTS AND
CC LINKS ACTIN FILAMENTS TO MEMBRANE GLYCOPROTEINS.
CC -!- SUBUNIT: HOMODIMER. INTERACTS WITH CVHSP.
CC -!- SUBCELLULAR LOCATION: PERIPHERAL CYTOPLASM.
CC -!- TISSUE SPECIFICITY: UBIQUITOUS.
CC -!- PTM: PHOSPHORYLATION EXTENT CHANGES IN RESPONSE TO CELL
CC ACTIVATION.
CC -!- SIMILARITY: THE ACTIN-BINDING DOMAIN IS OF A TYPE FOUND IN MANY
CC ACTIN-BINDING PROTEINS (SUCH AS ACTININ, DYSTROPHIN, FIMBRIN,
CC ABP-120, ABP-180, OR BETA-FODRIN).
CC -!- SIMILARITY: CONTAINS 2 CALPONIN-HOMOLOGY (CH) DOMAINS.
CC -!- SIMILARITY: CONTAINS 24 FILAMIN REPEATS.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: X53416; CAA37495.1; -
CC EMBL: L44140; AAA92644.1; -
CC EMBL: X70082; CAA49687.1; -
CC EMBL: X70085; CAA49690.1; -
CC PIR: A37098; A37098.
CC HSSP: P13466; 1KSR.
CC MIM: 300017; -
CC InterPro: IPR001589; Actinin_act_bind.
CC InterPro: IPR001715; Calponin_hom.
CC InterPro: IPR001298; Filamin.
CC Pfam: PF00307; CH; 2.
CC Pfam: PF00630; Filamin; 24.
CC SMART: SM00033; CH; 2.
CC PROSITE: PS00019; ACTININ_1; 1.
CC PROSITE: PS00020; ACTININ_2; 1.
CC PROSITE: PS00021; CH; 2.
CC PROSITE: PS0194; FILAMIN_REPEAT; 24.
CC Actin-binding; Phosphorylation; Repeat.
FT DOMAIN 1 274 ACTIN-BINDING (HEAD).
FT DOMAIN 43 149 CH 1.
FT REPEAT 166 266 CH 2.
FT REPEAT 276 374 FILAMIN 1.
FT REPEAT 376 474 FILAMIN 2.
FT REPEAT 475 570 FILAMIN 3.
FT REPEAT 571 663 FILAMIN 4.
FT REPEAT 667 763 FILAMIN 5.
FT REPEAT 764 866 FILAMIN 6.
FT REPEAT 867 965 FILAMIN 7.
FT REPEAT 966 1061 FILAMIN 8.
FT REPEAT 1062 1154 FILAMIN 9.
FT REPEAT 1155 1249 FILAMIN 10.
FT REPEAT 1250 1349 FILAMIN 11.
FT REPEAT 1350 1442 FILAMIN 12.
FT REPEAT 1443 1539 FILAMIN 13.
FT REPEAT 1540 1636 FILAMIN 14.
```

```

FT REPEAT 1649 1740 FILAMIN 15.
FT DOMAIN 1741 1778 HINGE 1.
FT REPEAT 1779 1860 FILAMIN 16.
FT REPEAT 1861 1950 FILAMIN 17.
FT REPEAT 1951 2039 FILAMIN 18.
FT REPEAT 2042 2131 FILAMIN 19.
FT REPEAT 2132 2230 FILAMIN 20.
FT REPEAT 2233 2325 FILAMIN 21.
FT REPEAT 2327 2420 FILAMIN 22.
FT REPEAT 2424 2516 FILAMIN 23.
FT DOMAIN 2517 2551 HINGE 2.
FT REPEAT 2552 2646 FILAMIN 24.
FT DOMAIN 2647 2677 SELF-ASSOCIATION SITE, TAIL.
FT SITE 1761 1762 CLEAVAGE (BY CALPAIN).
FT MOD_RES 1 1 BLOCKED.
FT CONFLICT 1772 1772 A -> G (IN REF. 3).
FT CONFLICT 2634 2634 H -> D (IN REF. 2).
SQ SEQUENCE 2647 AA; 280759 MW; 6C1A07041DFA3D42 CRC64;

Query Match 6.6%; Score 148.5; DB 1; Length 2647;
Best Local Similarity 21.8%; Pred. No. 1.2;
Matches 116; Conservative 66; Mismatches 190; Indels 159; Gaps 28;

QY 47 SEAYAYADLKKDNGEYTVVDADKG---YTLNFKFAGKEKTEPEP---KEEVYIKANLIY 100
DB 818 AEADIDFDIIRNDNFTVKYTPRGAGSYTIMVLEA-DQATPTSPTRVVRVPSHDASKVK 876
QY 101 ADGK-----TQTAFFKG-----TFEAT- AEARYADALKKDNGEYTV 138
DB 877 AEGPGLSRTGVELGKTHFTVNAKAAGKGLDVQFSLGTLKGAQVDRVDIIDHNDNTYTVK 936
QY 139 VADKGYT-----LNKIFAGKEKTEPEP-----KEE 163
DB 937 -----YTPVQGGPGVNVYGG-DPIPKSPFSAVSPSLDKIKVSLGEXVDVCKDOE 990
QY 164 VTIKANLIYADGK--TQTAFFKGTFPEATAEAYRYAD-----LLAKENGYTVVDADKGY 216
DB 991 FTVKSGAGGQGVASKIVGCPGSAAPVCPVEPLGADNSVVRFLPREEGPYEVEYTDGV 1050
QY 217 TL-NIKFAGKEKTEPEPKEEVTIKANLIYADG-----TKQTAFFKGTFPEA- --TA 263
DB 1051 PVPGPSFPLEAVAPTAKSK---VKA---FGPGLQGSAGSPARFTITDKAGTGGLGLTV 1104
QY 264 EAYRYADLLAKENGYTADL-----EDGGYTNIRFAGKKVDKPEP-----MDTYKLI 313
DB 1105 EGPCEAQLECLDNGDGTCSVSVYVTPPEPDYNINILFADTHIPGSPKHAHVPCFDASKVK 1164
QY 314 LNKGLTKGETTEA---VDAATA--EKVKQYANDNGVDGE---WTYODATKT----- 358
DB 1165 CSGPGLERATAGEVGQFQVDCSAGSAELTIEICSEAGLPAEYVIQDHGDGTHITYIPL 1224
QY 359 ----FTVTEK-----PEVIDASELTPAVTYTKLVINGKTLKE-----TTTKAVDAET 402
DB 1225 CPQAYVTYIKYGGQVPNPFPSKLQPEAVDTSVCVQCYGFCIGQGVFRATTEFSVDARA 1284
QY 403 AEKA-----FKQYANDNGVDGVW-----TYDDATKTFVT 432
DB 1285 LTQTGGPHVKARVANPSGNLITYVQDRG-DGMVKVEYTPYEGLHSVDVT 1334

RESULT 9
YFI6_YEAST STANDARD; PRT; 1233 AA.
AC P43597;
DT 01-NOV-1995 (Rel. 32, Created)
DT 01-NOV-1995 (Rel. 32, Last sequence update)
DE 01-NOV-1995 (Rel. 32, Last annotation update)
DE Hypothetical 137.7 kDa protein in UGSI-FAB1 intergenic region.
GN YFR016C.
OS Saccharomyces cerevisiae (Baker's yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Saccharomycetaceae; Saccharomyces.
OX NCBI_TaxID=4932;

```

```

RN SEQUENCE FROM N.A.
RP STRAIN=S288C / AB972;
RX MEDLINE=95400292; PubMed=7670463;
RA Murakami Y., Naitou M., Hagiwara H., Shibata T., Ozawa M.,
RA Sasanuma S.-I., Sasanuma M., Tsuchiya Y., Soeda E., Yokoyama K.,
RA Yamazaki M., Tashiro H., Eki T.;
RT "Analysis of the nucleotide sequence of chromosome VI from
RT Saccharomyces cerevisiae.";
RL Nat. Genet. 10:261-268(1995).
CC -!- SIMILARITY: SOME, TO MAMMALIAN NEUROFILAMENT TRIPLET M PROTEIN.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL: D50617; BAA09255.1; -
CC SGD: S0001912; YFR016C.
CC Hypothetical protein.
KW
SQ SEQUENCE 1233 AA; 137697 MW; C8A7CD2C6F0892F6 CRC64;

Query Match 6.6%; Score 148; DB 1; Length 1233;
Best Local Similarity 21.2%; Pred. No. 0.51;
Matches 99; Conservative 64; Mismatches 181; Indels 124; Gaps 16;

QY 2 VENKSETPTPTDSEEVYTIKANLIIFANGSTQTAFFKGTFPEKATS--EAYAYADTLKDD 59
DB 465 VEKEEEEEEEENSTFSKVKKENVGTGEQAVRNNEVSGTEESTSKGEIMGDEKQSE 524
QY 60 NGEYTVVDADKGYTLNFKFAGKEKTEPEPKEEVTIKANLIYAD-GKTQTAFFKG- ---- 113
DB 525 AGEKSIIEIGSANSKIS-----KDNLVLEDEAEAPTQENKPTVEVGE 569
QY 114 --FEATAEAYRYADALKKDNGEYTVVDADKGYTLNFKFAGKEKTEPEPKEEVTIKANLI 171
DB 570 IDPDAPRDDVIVEAVEKNIIPEDLEVAKEDQ-----EGQVKLDEPVK--AMKDDKI 621
QY 172 YADGKTQTAFFKGTFPEATAEAYRYADLLAKENGYTVVDADKGYTLNFKFAGKEKTEPEE 231
DB 622 AMRGAESTISEDMMKKQEGTAE-----LSNEKAKKVEDTARESAEGVE-VEKSKTPES 673
QY 232 PK-----EVTIKANLIYADCKTQTAFFK 255
DB 674 PKVVKRCTSGRPEDLOINERDPEVLKEDVRVPEDEVKPEIATTIENSEEDPKSRVQIS 733
QY 256 GTFAEATAEAYRYADLLAKENGYTADLEGGYTNIRFAGKKVDKPEPMDTYKLIILN 315
DB 734 TEQAETTK-----DMGDVGSTTSF-----KEEEKPKR----FEITQE 767
QY 316 GYTLKGETT-----TEAVDAATAEKFV-----KQYANDNGVDGWTYDDATK 357
DB 768 GSKITGKDTNHHGEATEAASNSKASDVGTAETIEPSSSVKDKDTEDAEVENSEKTE 827
QY 358 TPTVTEKPEVIDASELTPAVTYTKLVINGKTLKGETTTTKAVDAETAEK 405
DB 828 FIKVRAELENLDPK--EAEVTAELNKENEDVEVDTEEDA-EVENSEK 872

RESULT 10
DXT_STRDO STANDARD; PRT; 1337 AA.
AC P39653;
DT 01-FEB-1995 (Rel. 31, Created)
DT 01-FEB-1995 (Rel. 31, Last sequence update)
DT 15-JUL-1998 (Rel. 36, Last annotation update)
DE Dextranase precursor (EC 3.2.1.11) (Alpha-1,6-glucan-6-
DE glucanohydrolase).
GN DEX.
OS Streptococcus downei (Streptococcus sobrinus).

```

OG Plasmid pYA902.
 OC Bacteria; Firmicutes; Bacillus/Clostridium group; Streptococcaceae;
 OC Streptococcus.
 OX NCBI_TaxID=1317;
 RN [1]
 RP SEQUENCE FROM N.A., AND SEQUENCE OF 31-36.
 RC STRAIN=6715 / UAB66;
 RX MEDLINE=94292401; PubMed=80211165;
 RA Wanda S.-Y., Curtiss R. III;
 RT "Purification and characterization of Streptococcus sobrinus
 RT dextranase produced in recombinant Escherichia coli and sequence
 RT analysis of the dextranase gene."
 RL J. Bacteriol. 176:3839-3850(1994).
 CC -!- FUNCTION: MAY PLAY A ROLE IN SUCROSE-INDEPENDENT ADHERENCE TO THE
 CC PELICLE-COATED TOOTH SURFACE. THE ACTIVITY OF THIS ENZYME IS
 CC OPTIMAL AT PH 5.3 AND AT 39 DEGREES CELSIUS.
 CC -!- CATALYTIC ACTIVITY: Endohydrolysis of 1,6-alpha-D-glucosidic
 CC linkages in dextran.
 CC -!- SUBUNIT: HOMODIMER.
 CC -!- SUBCELLULAR LOCATION: Type I membrane protein. Cell wall.
 CC -!- SIMILARITY: TO OTHER STREPTOCOCCAL AND STAPHYLOCOCCAL PROTEINS
 CC IN THE REGION OF THE MEMBRANE ANCHOR.
 CC -!- SIMILARITY: BELONGS TO FAMILY 66 OF GLYCOSYL HYDROLASES.
 CC -!- CAUTION: IT IS UNCERTAIN WHETHER MET-1 OR MET-5 IS THE INITIATOR.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC
 DR EMBL: M96978; AAA21772.1;
 DR InterPro: IPR001899; Gram_pos_anchor.
 DR Pfam: PF00746; Gram_pos_anchor; 1.
 DR PROSITE: PS00343; Gram_pos_anchoring; 1.
 KW Hydrolase; Glycosidase; Signal; Transmembrane; Repeat; Plasmid.
 FT SIGNAL 1 30
 FT CHAIN 31 1337 DEXTRANASE.
 FT DOMAIN 31 1313 EXTRACELLULAR (BY SIMILARITY).
 FT TRANSMEM 1314 1332 MEMBRANE ANCHOR (BY SIMILARITY).
 FT DOMAIN 1333 1337 CYTOPLASMIC (BY SIMILARITY).
 FT DOMAIN 1305 1310 CONSERVED IN GRAM-POSITIVE COCCI SURFACE
 FT PROTEINS.
 SQ SEQUENCE 1337 AA; 143298 MW; B494275A77A2E3D0 CRC64;
 Query Match 6.6%; Score 146.5; DB 1; Length 1337;
 Best Local Similarity 20.9%; Pred. No. 0.68;
 Matches 110; Conservative 64; Mismatches 181; Indels 171; Gaps 26;
 QY 6 EETPETPET-DSEEEVTIKANLI--FANGSTQ---TAEEKGFTEKATSEAYAYADTLKK 58
 Db 100 EQTP--PETSASAPATTSADSVKEVQAQDATQNSTGNGPVIRATSAQVATRSVWS 157
 QY 59 DNGEYTVDV-ADK-----GYTLNIKFPAGKEKTPPEEKVEVTIKANLIYADGKTQTAEKFG 112
 Db 158 QSGDAIVLSADKASRYQGEDVNLSDVDFKNTTDE--QDVTVADVYYIDNKLGT----- 210
 QY 113 TEEATAEAYRYRADALKKNGEYTVDVADKGYTLNKFAGKEKTPP---EPKEEVTIKAN 169
 Db 211 -----YKFSKHLK-----AGEGYKMQ---SGDLKIPASQFENNHYLLKVR 248
 QY 170 LIYADGKTQTAEKGFTEKATEAY-RYA-----DLLAKENGKVTVDVAD-KGYT 217
 Db 249 VRDADNLTUSEVKAIAVEDSWTKFPYGIYVGSQDNTNLSLKADRYAEIEKMKNN 308
 QY 218 LNIKF---AGKEKTPPEEKVEYTIKA-----NL 242
 Db 309 INSIFYDYVYKATNPFPSPDEATFKQDMWNWSSSEIDTQAVKDYNQVHGGNAVAMLYNM 368
 QY 243 IYADGKTQTAEF-----KGTFA-----EATAEAYRYADLL 272

RESULT 11

WAPA_BACSU STANDARD; PRT: 2334 AA.
 ID Q07833;
 AC DT 01-OCT-1994 (Rel. 30, Created)
 DT 01-OCT-1994 (Rel. 30, Last sequence update)
 DT 01-MAR-2002 (Rel. 41, Last annotation update)
 DE Wall-associated protein precursor.
 GN WAPA OR NI7G.
 OS Bacillus subtilis.
 OC Bacteria; Firmicutes; Bacillus/Clostridium group;
 OC Bacillus/Staphylococcus group; Bacillus.
 OX NCBI_TaxID=1423;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=168;
 RX MEDLINE=93302506; PubMed=8316082;
 RA Foster S. J.;
 RT Molecular analysis of three major wall-associated proteins of
 RT Bacillus subtilis 168: evidence for processing of the product of a
 RT gene encoding a 258 kDa precursor two-domain ligand-binding
 RT protein.";
 RL Mol. Microbiol. 8:299-310(1993).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=168 / BGSC1A1;
 RX MEDLINE=95219088; PubMed=7704263;
 RA Yoshida K.-I., Sano H., Seki S., Oda M., Fujimura M., Fujita Y.;
 RT "Cloning and sequencing of a 29 kb region of the Bacillus subtilis
 RT genome containing the hut and wapa loci";
 RL Microbiology 141:337-343(1995).
 RN [3]
 RP SEQUENCE FROM N.A.
 RC STRAIN=168 / BGSC1A1;
 RX MEDLINE=97124196; PubMed=8969509;
 RA Yoshida K.-I., Shindo K., Sano H., Seki S., Fujimura M., Yanai N.,
 RA Miwa Y., Fujita Y.;
 RT "Sequencing of a 65 kb region of the Bacillus subtilis genome
 RT containing the lic and cel loci, and creation of a 177 kb contig
 RT covering the gnt-sacxy region.";
 RL Microbiology 142:3113-3123(1996).
 CC -!- FUNCTION: STILL UNKNOWN. NOT INVOLVED IN CELL MEMBRANE METABOLISM,
 CC MOTILITY, SECRETION OR DIFFERENTIATION.
 CC -!- SUBCELLULAR LOCATION: MEMBRANE-ASSOCIATED. MAY BE RELEASED
 CC INTO THE MEDIUM.
 CC -!- DOMAIN: HAS TWO LIGAND-BINDING DOMAINS; THE N-TERMINUS, HAS THREE
 CC 101 AA REPEATS WHICH ARE RESPONSIBLE FOR CELL WALL BINDING; THE
 CC C-TERMINUS CONSISTS OF TWO BLOCKS OF RESIDUES WITH A CONSERVED
 CC MOTIF REPEATED 31 TIMES.
 CC -!- SIMILARITY: THE REPEATED MOTIF IN THE C-TERMINUS HAS SOME
 CC SIMILARITY TO THE REPEAT IN E.COLI RHS GROUP OF PROTEINS (RHS-A-D).
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC

```
CC or send an email to license@sib.ch).
CC -----
DR EMBL; L05634; AAA22883.1; -
DR EMBL; D31856; BAA06656.1; -
DR EMBL; D29985; BAA06260.1; -
DR EMBL; D83026; BAA11683.1; -
DR EMBL; Z99124; CAB15959.1; -
DR PIR; S32920; S32920.
DR Subtilist; BG10797; wapA.
DR InterPro; IPR003305; CBD_6.
DR Pfam; PF02018; CBD_6; 1.
KW Cell wall; Repeat; Signal; Complete proteome.
FT SIGNAL 1 28 OR 32 (POTENTIAL).
FT CHAIN 29 2334 WALL-ASSOCIATED PROTEIN.
FT DOMAIN 504 869 3 X 101 AA APPROXIMATE TANDEM REPEATS.
FT REPEAT 504 605 1-1.
FT REPEAT 636 736 1-2.
FT REPEAT 769 869 1-3.
FT DOMAIN 1021 2139 31 X 21 AA APPROXIMATE TANDEM REPEATS OF
X(4)-G-X(4)-[YF]-X-D-X(2)-G-X(4).
FT REPEAT 1021 1040 2-1.
FT REPEAT 1042 1061 2-2.
FT REPEAT 1063 1082 2-3.
FT REPEAT 1083 1102 2-4.
FT REPEAT 1109 1128 2-5.
FT REPEAT 1129 1148 2-6.
FT REPEAT 1150 1169 2-7.
FT REPEAT 1174 1193 2-8.
FT REPEAT 1199 1218 2-9.
FT REPEAT 1219 1238 2-10.
FT REPEAT 1646 1665 2-11.
FT REPEAT 1667 1686 2-12.
FT REPEAT 1690 1709 2-13.
FT REPEAT 1711 1730 2-14.
FT REPEAT 1732 1751 2-15.
FT REPEAT 1753 1772 2-16.
FT REPEAT 1795 1814 2-17.
FT REPEAT 1820 1839 2-18.
FT REPEAT 1840 1859 2-19.
FT REPEAT 1861 1880 2-20.
FT REPEAT 1887 1906 2-21.
FT REPEAT 1908 1927 2-22.
FT REPEAT 1929 1948 2-23.
FT REPEAT 1969 1982 2-24.
FT REPEAT 1983 2002 2-25.
FT REPEAT 2008 2027 2-26.
FT REPEAT 2028 2047 2-27.
FT REPEAT 2051 2070 2-28.
FT REPEAT 2071 2090 2-29.
FT REPEAT 2093 2112 2-30.
FT REPEAT 2120 2139 2-31.
SQ SEQUENCE 2334 AA; 258329 MW; B75138CCDD278BAA3 CRC64;

Query Match 6.4%; Score 144; DB 1; Length 2334;
Best Local Similarity 21.1%; Pred. No. 1.8;
Matches 118; Conservative 70; Mismatches 214; Indels 156; Gaps 27;

QY 14 TDSEEVTTKANLIFANGSTQTAERKGTTEKATSEAYAYADTLTKDNGEYTVDA---DK 70
DB 1534 TD-EKDYSLQANVTYADGS--TGIYNAKTPSGTQENRAAVVPIKTPINKVDISILFQK 1590
QY 71 GYTLNFKF-----AGKEKTPPEEKVTTKANLIYADGKTTQTAERFKG 112
DB 1591 SARTGVWFDDIRLEIGSLTLTKSYDSNGVVTKEEDLGYATSTDYDETKKTSSETDAKG 1650
QY 113 TFPEATAYARYADAL-----KKDNGEYTVDV---ADKGYTLNFKPAGKE 154
DB 1651 ---EKTYYDQDLQTLNLTSLNGTSILHSYDKEGNEVSKTRAGADQYKFEYDVMGKL 1707
QY 155 KTPPEEP-----KEVTTKANL---IYADGKTQTAERFKGTFEATAPAY---RYADLLAK 202
DB 1708 VKTTDPLGNVLASEYDANSLTKTISPNGNEVSLSDYGT-DRVKSKSYNGTEKIPTYDK 1766
```


QY 346 VDGEWTVDDATKFTTVEKPEVIDASELTPAVTYK--LVINGKTLKGET-----TT 395
Db 657 -----DATAQVTVONNSVITVGQAKAGETKYVTVVLDGKLITTHSFKVVDTPAPTA 708
QY 396 KAVDAETAFAKQYA-----NDNGVDGV 419
Db 709 KGLAVEFTSTSLKEVAPNADLKAALLNLSVDGV 742

RESULT 14
NEM_RAT
ID NEM_RAT STANDARD; PRT; 845 AA.
AC P12839; Q63370;
DT 01-OCT-1989 (Rel. 12, Created)
DT 01-MAY-1992 (Rel. 22, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Neurofilament triplet M protein (160 kDa neurofilament protein)
DE (Neurofilament medium polypeptide) (NF-M).
GN NEF3 OR NEFM OR NEFM.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_TaxID=10116;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=87282618; PubMed=2441012;
RA Napolitano E.W., Chin S.S.M., Colman D.R., Liem R.K.H.;
RT "Complete amino acid sequence and in vitro expression of rat NF-M,
the middle molecular weight neurofilament protein.";
RL J. Neurosci. 7:2590-2599(1987).
RN [2]
RP SEQUENCE FROM N.A.
RX STRAIN=WISTAR;
RX MEDLINE=92332596; PubMed=1321159;
RA Kelly B.M., Gillespie C.S., Sherman D.L., Brophy P.J.;
RT "Schwann cells of the myelin-forming phenotype express neurofilament
protein NF-M.";
RL J. Cell Biol. 118:397-410(1992).
RN [3]
RP PHOSPHORYLATION SITES, AND REVISION TO 500.
RX MEDLINE=92165797; PubMed=1537832;
RA Xu Z.-S., Liu W.-S., Willard M.B.;
RT "Identification of six phosphorylation sites in the COOH-terminal
tail region of the rat neurofilament protein M.";
RL J. Biol. Chem. 267:4467-4471(1992).
RN [4]
RP CARBOHYDRATE-LINKAGE SITES.
RX MEDLINE=93346421; PubMed=8344946;
RA Dong D.L.-Y., Xu Z.-S., Chevrier M.R., Cotter R.J., Cleveland D.W.,
RA Hart G.W.;
RT "Glycosylation of mammalian neurofilaments. Localization of multiple
O-linked N-acetylglucosamine moieties on neurofilament polypeptides
L and M.";
RL J. Biol. Chem. 268:16679-16687(1993).
CC -1- FUNCTION: NEUROFILAMENTS USUALLY CONTAIN THREE IF PROTEINS: L, M,
AND H WHICH ARE INVOLVED IN THE MAINTENANCE OF NEURONAL CALIBER.
CC -1- PTM: THERE ARE A NUMBER OF REPEATS OF THE TRIPEPTIDE K-S-P, NFM IS
PHOSPHORYLATED ON A NUMBER OF THE SERINES IN THIS MOTIF. IT IS
THOUGHT THAT PHOSPHORYLATION OF NEF RESULTS IN THE FORMATION OF
INTERFILAMENT CROSS BRIDGES THAT ARE IMPORTANT IN THE MAINTENANCE
OF AXONAL CALIBER.
CC -1- PTM: PHOSPHORYLATION SEEMS TO PLAY A MAJOR ROLE IN THE FUNCTIONING
OF THE LARGER NEUROFILAMENT POLYPEPTIDES (NF-M AND NF-H). THE
LEVELS OF PHOSPHORYLATION BEING ALTERED DEVELOPMENTALLY AND
COINCIDENT WITH A CHANGE IN THE NEUROFILAMENT FUNCTION.
CC -1- SIMILARITY: BELONGS TO THE INTERMEDIATE FILAMENT FAMILY.
CC
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>)

CC or send an email to license@isb-sib.ch.
CC
DR EMBL; M18628; AAA41696.1; -
DR EMBL; Z12152; CAA78136.1; -
DR GlycoSuiteDB; P12839; -
DR InterPro; IPR001664; IF.
DR Pfam; PF00038; filament; 1.
DR PROSITE; PS00226; IF; 1.
KW Intermediate filament; Coiled coil; Neurone; Phosphorylation;
FT INIT_MET 0 0 BY SIMILARITY.
FT DOMAIN 1 103 HEAD.
FT DOMAIN 103 410 ROD.
FT DOMAIN 411 844 TAIL.
FT DOMAIN 103 134 COIL 1A.
FT DOMAIN 135 147 LINKER 1.
FT DOMAIN 148 246 COIL 1B.
FT DOMAIN 247 263 LINKER 12.
FT DOMAIN 264 285 COIL 2A.
FT DOMAIN 286 289 LINKER 2.
FT DOMAIN 290 410 COIL 2B.
FT CARBOHYD 47 O-LINKED (GLCNAC).
FT CARBOHYD 47 /FTID-CAR_000130.
FT CARBOHYD 430 O-LINKED (GLCNAC).
FT MOD_RES 502 /FTID-CAR_000131.
FT MOD_RES 506 PHOSPHORYLATION.
FT MOD_RES 536 PHOSPHORYLATION.
FT MOD_RES 603 PHOSPHORYLATION.
FT MOD_RES 608 PHOSPHORYLATION.
FT MOD_RES 666 PHOSPHORYLATION.
FT CONFLICT 17 17 MISSING (IN REF. 2).
FT CONFLICT 21 21 R -> P (IN REF. 2).
FT CONFLICT 204 204 V -> L (IN REF. 2).
FT CONFLICT 500 500 MISSING (IN REF. 1).
SQ SEQUENCE 845 AA; 95660 MW; 316C41655B11197D CRC64;
Query Match 6.3%; Score 141.5; DB 1; Length 845;
Best Local Similarity 23.4%; Pred. No. 0.73;
Matches 92; Conservative 51; Mismatches 154; Indels 97; Gaps 16;
QY 5 KEETPTPTDSEEEVITKANIFANGSTQTAEFGTFEKATSEAYAYADTLKKONGEY 64
Db 517 KBEEREGQEEEEEDEGVKSDQAEEGS-----EKEGSEKDEGEQEEGETEAEGERA 572
QY 65 V-----DVADKGYTLNIFAGKEK-----TPEEPKEEVIKANLIYADGKTQTAE 111
Db 573 EAKEKKTEGVVEEMAIKEEIKVEKPEKAKSPVKSPVEEVKPKPEA---KAGDKQKEE 630
QY 112 GTFE--EATAEAYRYADALKDNGEYTVADVADKGYTLNIFAGKEKTPPEPEEVIKAN 169
Db 631 KYEEKKEVAKESPKKEKVEKE--EKPKDVPDK-----KKAESPVKKEAVE-E 675
QY 170 LIYADGKTQTAEFGTFEETAEAYRYADLLAKENGKYYTVADVADKGYTLNIFAGKEKTP 229
Db 676 MITITKSVKVSLEKDTKEEKPOQKEKVE--KAEEEGGSEEEVGD-----KSP 721
QY 230 EE-PKEEVIKANLIYADGKTQTAEFGTFEETAEAYRYADLLAKENGKYYTADLEDGY 288
Db 722 QESKEDIAINGEVGKEEEEOFTOEKSGQE-----EERGVVTVNGLD----- 764
QY 289 TINIRFAGKKVKEPEEPMDTYKLLINGKTLKGETTTTAEVDAATAEKVKFYQYANDNGVDG 348
Db 765 -----VSPAEEKKGEDSDDKVVV-----TKKVEKITSE-----GGDG 797
QY 349 EWYDDATKFTTVEKPEVIDASELTPAVTYK 382
Db 798 ATKY--ITKSVTVTKVEEHEETFEKVLSTKKV 829
RESULT 15
ID MAPB_RAT STANDARD; PRT; 2459 AA.

THIS PAGE BLANK (USPTO)

RESULT 2
A42808
Ig light chain-binding protein precursor - Peptostreptococcus magnus
N:Alternate names: protein L
C:Species: Peptostreptococcus magnus
C:Date: 19-May-1994 #sequence_revision 19-May-1994 #text_change 07-May-1999
C:Accession: A42808; A41493
R:Kastern, W.; Sjoerbring, U.; Bjoerck, L.
J. Biol. Chem. 267, 12820-12825, 1992, 1990
A:title: Structure of peptostreptococcal protein L and identification of a repeated immunoglobulin domain
A:Reference number: A42808; MUID:92316971
A:Accession: A42808
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-719 <KAS>
A:Cross-references: GB:M86697
R:Kastern, W.; Holst, E.; Nielsen, E.; Sjoerbring, U.; Bjoerck, L.
Infect. Immun. 58, 1217-1222, 1990
A:title: Protein L, a bacterial immunoglobulin-binding protein and possible virulence determinant
A:Reference number: A41493; MUID:90215984
A:Accession: A41493
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 202-275 <KA2>
C:Keywords: immunoglobulin

Query Match 71.0%; Score 1587.5; DB 2; Length 719;
Best Local Similarity 76.3%; Pred. No. 3.2e-76;
Matches 334; Conservative 13; Mismatches 56; Indels 35; Gaps 6;

QY 2 VENKEETPETPTDSEEVTIKANLIFANGSTQTAEPKGFATSEAYAYADTLKKDNG 61
DB 95 VENKEETPETPTDSEEVTIKANLIFANGSTQTAEPKGFATSEAYAYADTLKKDNG 194

QY 62 EYTVADVADKGYTLNIFAGKEKTPPEPKEEVTIKANLIYADGKTQTAEPKGFEEATAEA 121
DB 155 EYTVADVADKGYTLNIFAGKEKTPPEPKEEVTIKANLIYADGKTQTAEPKGFEEATAEA 214

QY 122 YRVADALKKDNGETVVDVADKGYTLNIFAGKEKTPPEPKEEVTIKANLIYADGKTQTA 191
DB 215 YRVADALKKDNGETVVDVADKGYTLNIFAGKEKTPPEPKEEVTIKANLIYADGKTQTA 274

QY 182 FKGTFFEATAEAYRYADLLAKENGYTVVDVADKGYTLNIFAGKEKTPPEPKEEVTIKAN 241
DB 275 FKGTFFEATAEAYRYADLLAKENGYTVVDVADKGYTLNIFAGKEKTPPEPKEEVTIKAN 334

QY 242 LIYADGKTQTAEPKGFATSEAYAYADLLAKENGYTADLEDGGYTTINIRFAGKKVDE 301
DB 335 LIYADGKTQTAEPKGFATSEAYAYADLLAKENGYTADLEDGGYTTINIRFAGKKVDE 394

QY 302 KPPEPMD-TYK--LILNGKTLKGETTEAYDAATAEAKVFKQYANDNGVD-GEWYDDATK 357
DB 395 KPPEKQVTTIKENIYFDGTVQVATFGTFAEATAEAY--RYADLLSKEHGKTYADLEDG 452

QY 358 TFFVT-----EKPEVIDASELTPAVTTKLVNGTKLGETTTKAVDAETA 404
DB 453 GYINIRFAGKEPPEETPEPQVD-----GYASVEAEAAKALKND 496

QY 405 KAFKQYANDNGVGVWY 422
DB 497 DVNKSVTIRQAGDGRYYY 514

RESULT 3
S54396
protein L precursor - Peptostreptococcus magnus (strain 3316)
C:Species: Peptostreptococcus magnus
A:Variety: strain 3316
C:Date: 27-Oct-1996 #sequence_revision 13-Mar-1997 #text_change 15-Oct-1999
C:Accession: S54396
R:Murphy, J.P.; Duggleby, C.J.; Atkinson, M.A.; Trowern, A.R.; Atkinson, T.; Goward, C.H.

Mol. Microbiol. 12, 911-920, 1994
A:title: The functional units of a peptostreptococcal protein L.
A:Reference number: S54396; MUID:95020613
A:Accession: S54396
A:Status: preliminary; nucleic acid sequence not shown
A:Molecule type: DNA
A:Residues: 1-992 <MUR>
A:Cross-references: EMBL:L04466; NID:g150673; PIDN:AAA67503.1; PID:g150674

Query Match 56.5%; Score 1263.5; DB 2; Length 992;
Best Local Similarity 58.8%; Pred. No. 4.9e-59;
Matches 281; Conservative 36; Mismatches 90; Indels 71; Gaps 12;

QY 2 VENKEETPETPTDSEEVTIKANLIFANGSTQTAEPKGFATSEAYAYADTLKKDNG 61
DB 241 MERKLEKETPE--PEEVTIKANLIFADGSTONAEKGFATSEAYAYADTLKKDNG 298

QY 62 EYTVADVADKGYTLNIFAGKEKTPPEPKEEVTIKANLIYADGKTQTAEPKGFEEATAEA 121
DB 299 EYTVADVADKGYTLNIFAGKEKTPPEPKEEVTIKANLIFADGKTQTAEPKGFEEATAEA 358

QY 122 YRVADALKKDNGETVVDVADKGYTLNIFAGKEKTPPEPKEEVTIKANLIYADGKTQ 179
DB 359 YAYADLLAKENGYTADLEDGGNTINIFAGKETPETPEPKEEVTIKNLIYADGKIQ 418

QY 180 AEPKGFEEATAEAYRYADLLAKENGYTVVDVADKGYTLNIFAGKEKTPPEPKEEVT 237
DB 419 AEPKGFEEATAEAYRYADLLAKENGYTADLEDGGNTINIFAGKETPETPEPKEEVT 478

QY 238 IKANLIYADGKTQTAEPKGFATSEAYAYADLLAKENGYTADLEDGGYTTINIRFAGK 297
DB 479 IKNLIYADGKTQTAEPKGFATSEAYAYADLLAKENGYTADLEDGGYTTINIRFAGK 538

QY 298 K-----VDE-----KPPEP-----DPY-KLILNGKTLG----- 321
DB 539 EQGENPGTITDEWLKNAKEAEIKELKZAGTSDLYFLINKAKTVEGVKALNKLKA 598

QY 322 -----ET-----TTEAVDAATAEAKVFKQYAND-----NGVGEWYDDATFTFVTEK 364
DB 599 HAGEETPELKDGATYEEAEAAKALKNDVNNAYEIVOGADGRYYY--VLKIEVADEE 656

QY 365 PEVIDASELTPAVTTKLVNGTKLGETTTKAVDAETAEAKKQYANDNGVGVWY 422
DB 657 EPGEDTPEVQEGYATYE-----EAEAAKALKEDKVNNAVYEVVQADGRYYY 704

RESULT 4
S00128
protein G precursor - Streptococcus sp. (Streptococcus G148)
N:Alternate names: albumin-binding protein; cell wall-bound protein
C:Species: Streptococcus sp.
A:Variety: Streptococcus G148
C:Date: 30-Jun-1989 #sequence_revision 30-Jun-1989 #text_change 19-May-2000
C:Accession: S00128; A27604; A26314
R:Olsson, A.; Eliasson, M.; Guss, B.; Nilsson, B.; Hellman, U.; Lindberg, M.; Uhlen, Eur. J. Biochem. 168, 319-324, 1987
A:title: Structure and evolution of the repetitive gene encoding streptococcal protein G
A:Reference number: S00128; MUID:88029445
A:Accession: S00128
A:Molecule type: DNA
A:Residues: 1-593 <OLS>
A:Cross-references: EMBL:X06173; NID:g47084; PIDN:CAA29540.1; PID:g47085
A:Note: the source is designated as Streptococcus G148
A:Note: part of this sequence, including the amino end of the mature protein, was con R:Sjoerbring, U.; Falkenberg, C.; Nielsen, E.; Akerstroem, B.; Bjoerck, L. J. Immunol. 140, 1595-1599, 1988
A:title: Isolation and characterization of a 14-kDa albumin-binding fragment of strep A:Reference number: A27604; MUID:88154455
A:Accession: A27604
A:Molecule type: protein
A:Residues: 62-101 <SJO>
R:Guss, B.; Eliasson, M.; Olsson, A.; Uhlen, M.; Frej, A.K.; Jornvall, H.; Flock, J.I. EMBO J. 5, 1567-1575, 1986

A;Title: Structure of the IgG-binding regions of streptococcal protein G.
A;Reference number: A26314; MUID:86300657
A;Accession: A26314
A;Molecule type: DNA
A;Residues: 114-593 <GUS>
A;Cross-references: GB:X04015; NID:g47071; PIDN:CAA27638.1; PID:g47072
C;Function:
C;Description: it is part of the cell wall structure of group G streptococci and is cov

C;Superfamily: M5 protein
C;Keywords: duplication; membrane protein
F;1-33/Domain: signal sequence #status predicted <SIG>
F;34-593/Product: protein G #status experimental <MAT>
F;34-116/Domain: alanine-rich <ALA>
F;117-290/Domain: AB duplication <DUPL>
F;117-140/Region: A repeat
F;141-191/Region: B repeat
F;192-215/Region: A repeat
F;216-266/Region: B repeat
F;267-290/Region: A repeat
F;303-497/Domain: IgG binding <IGB>
F;303-357/Region: C repeat
F;358-372/Region: D repeat
F;373-427/Region: C repeat
F;428-442/Region: D repeat
F;443-497/Region: C repeat
F;498-567/Domain: proline-rich <PRO>
F;568-593/Domain: carboxyl-terminal <CTD>

Query Match 31.7%; Score 708; DB 2; Length 593;
Best Local Similarity 48.3%; Pred. No. 3.7e-30;
Matches 205; Conservative 28; Mismatches 131; Indels 60; Gaps 17;

QY 45 ATSEAYAYADTLKKDNGEYVDVADKGYTLNIRFAGKEKTPPEPKKEVTKANLIYA-DG 103
DB 101 ADALAKADALKEFN-KYGVSDYKRLNNAKTVGKDLQAOVSESARKARISEATDG 159
QY 104 -----KTQTAEFKGTFFEEATAEAYRYADALKKONGE---YTVDVADKGYTLNIRFAGKEK 155
DB 160 LSDFLKSQTP-----AEDTVKSTELAEAKVLANRELDKYGVSDYHKNLNNAKTVEGCVK 213
QY 156 TPEPKKEVTKANLIYA-DG-----KTQTAEFKGTFFEEATAEAYRYAD---LLAKENGK 206
DB 214 DLQAOVSESARKARISEATDGLSDFLKSQTP-----AEDTVKSTELAEAKVLANRELDK 267
QY 207 YTVDVADKGYTLNIRFAGKEKTPPE-----PKKEVTKANLIYADKGTQTAEFKGTFAE 260
DB 268 YGVSDYKRLNNAKTVGKDLQAOVSESARKARISEATDGLSDFLKSQTP-----LILNGKTLKE---TTTE 320
QY 261 A--TAEAYRYADLLAKENGKGTADLEDGGYTINIRFAGKVKDEKPE-----EPMDTY 310
DB 321 AVDAATAEKVKQYANDNG-----VDGEWYDDATKTFVTEKPEVIDASELTPAVTTY 374
QY 311 KLINGKTLKGETTEAVDAATAEKVKQYANDNGVGEWYDDATKTFVTEKPEVIDA 370
DB 375 KLINGKTLKGETTEAVDAATAEKVKQYANDNGVGEWYDDATKTFVTEKPEVIDA 434
QY 371 SELTPAVTTYKLVTNGTKLGETTTKAVDAETAETAEKAFKQYANDNGVDGVTYDDATKTF 430
DB 435 SELTPAVTTYKLVTNGTKLGETTTKAVDAETAETAEKAFKQYANDNGVDGVTYDDATKTF 494
QY 431 VTEM 434
DB 495 VTEM 498

RESULT 5
S42574
streptococcal surface protein - Streptococcus dysgalactiae
C;Species: Streptococcus dysgalactiae
C;Date: 13-Jan-1995 #sequence_revision 13-Jan-1995 #text_change 19-May-2000
A;Reference number: S42574
R;Jonsson, H.; Mueller, H.P.
Eur. J. Biochem. 220, 819-826, 1994

A;Title: The type-III Fc receptor from Streptococcus dysgalactiae is also an alpha(2)
A;Reference number: S42574; MUID:94192673
A;Accession: S42574
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-664 <ON>
A;Cross-references: EMBL:Z29666; NID:g470220; PIDN:CAA82764.1; PID:g470221
C;Superfamily: M5 protein

Query Match 31.6%; Score 706.5; DB 2; Length 664;
Best Local Similarity 43.9%; Pred. No. 5.2e-30;
Matches 192; Conservative 51; Mismatches 143; Indels 51; Gaps 17;
QY 24 ANLIIFANGSTQTAEFKGTFFKATSEAYAYADTLK-KDNGEYTVDVADKGYTLNIRFAGKEK 79
DB 148 ANEIVNNSDAYTAESIOPLYKLINDAY--DVLESKDYSKY--DSQDKVNNLADQLRDVA 202
QY 80 -----GKEKTPPEPKKEVTKANLIYADKGTQTAEFKGTFFEEATAE---YRYA 125
DB 203 QAVQLEAPTVIDAPELTPALTTVTKLVVKGNTF--SGETTTK-----AIDTATAEKEFKQYA 256
QY 126 DALKKDNGEYTVDVADKGYTLNIRFAGKEKTPPEPKKEVTKANLIYADKGTQTAEFK 182
DB 257 TANNVD-GEWSYDDATKTFVTEKPAVIDAPELTPALTTVTKLVVKGNTF--SGETTT--- 310
QY 183 KGTFFEEATAEAYRYADLLAKENGKGYTVDVADKGYTLNIRFAGKEKTPPEPKKEVTK 239
DB 311 KAVDAETAETAEKAFKQYATANNVDSYDDATKTFVTEKPAVIDAPELTPALTTVTKLV 370
QY 240 ANLIYADKGTQTAEFKGTFFKATSEAYAYADLLAKENGKGTADLEDGGYTINIRFAGK 299
DB 371 GNTF--SGETTTKRAIDAATAEKEFKQYATANGV--DGWSYDDATKTFVTEKPA--V 422
QY 300 DEKPE--EPMDTYKLINGKTLKGETTEAVDAATAEKVKQYANDNGVGEWYDDATK 357
DB 423 IDAPELTPALTTVTKLVVKGNTFSGETTTKAVDAETAETAEKAFKQYANENGVSYSDDATK 482
QY 358 TFFVTEKPEVIDASELTPAVTTYKLVTNGTKLGETTTTAVDAETAETAEKAFKQYANDNGVD 417
DB 483 TFFVTEKPAVIDAPELTPALTTVTKLVVKGNTFSGETTTKAVDAETAETAEKAFKQYANENGVD 542
QY 418 GVWYDDATKTFVTEM 434
DB 543 GVWYDDATKTFVTEM 559

RESULT 6
A24496
IgG-binding protein - Streptococcus sp. (group G)
C;Species: Streptococcus sp.
C;Date: 17-Sep-1987 #sequence_revision 17-Sep-1987 #text_change 17-Mar-2000
C;Accession: A24496; A39041
R;Fahnestock, S.R.; Alexander, P.; Nagle, J.; Filpula, D.
J. Bacteriol. 167, 870-880, 1986
A;Title: Gene for an immunoglobulin-binding protein from a group G Streptococcus.
A;Reference number: A24496; MUID:86304178
A;Accession: A24496
A;Molecule type: DNA
A;Residues: 1-448 <FAH>
A;Cross-references: GB:M13825; NID:g153822; PIDN:AAA03664.1; PID:g153823
R;Sjoeborg, U.; Bjoerck, L.; Kastern, W.
J. Biol. Chem. 266, 399-405, 1991
A;Title: Streptococcal protein G. Gene structure and protein binding properties.
A;Reference number: A39041; MUID:91093154
A;Accession: A39041
A;Status: preliminary
A;Molecule type: protein
A;Residues: 34-42, N', 45-48; 62-76; 186-200 <SJO>
C;Genetics:
A;Gene: spg
C;Superfamily: M5 protein
C;Keywords: transmembrane protein

```

Query Match              30.6%; Score 683; DB 2; Length 448;
Best Local Similarity    53.2%; Pred. No. 5.4e-29;
Matches 176; Conservative 15; Mismatches 92; Indels 48; Gaps

Qy 130 KNGEYTVDVADKGYTLNKFAGKEKTPPEPKPEVITKANLIYAD-GKQTQAEFKGTFEE 188
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 45 RNGGELT-----NLLGNSETTLALRNESATADLTAAAVADVTA AAAAENAGA 92

Qy 189 ATAPAYRYADLLA-----KENGKTYVDVADKGYTLNKFAGKEKTPPEPKPEVITKAN 241
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 93 AAWEAAAAADALAKAKADALKEFKGVSDYYKNLNNAKTVEGIDKDLQAQVVESAKKAR 152
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Qy 242 LIVA-DG---KTQAEFKGTFAEATAEAYRYAD--LLAKENGKYTADLEDGGYITNI 292
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 153 ISEATDGLSDFLKSQTP-----AEDTVKSIELAEAKVLANRELDKIGV----SDYHKML 202

Qy 293 RFAGKKYDEKPE-----EPMDTYKLLINGKTLKGTTTTAEVDAATAEKVQKYAND 343
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 203 INNAKTVEGVKELIDEILAAPKTDYTKLLINGKTLKGTTTTAEVDAATAEKVQKYAND 262

Qy 344 NGVDGEWYTDATKTFVTVEKPEVIDASELTPAVTTYKLVINGKTLKGETTTKAVDAETA 403
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 263 NGVDGEWYTDATKTFVTVEKPEVIDASELTPAVTTYKLVINGKTLKGETTTKAVDAETA 322

Qy 404 EKAFQKYANDNGVDGWYTDATKTFVTTEM 434
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 323 EKAFQKYANDNGVDGWYTDATKTFVTTEM 353

RESULT 7
S55890
C:plasma protein receptor MAG precursor - Streptococcus dysgalactiae
C:Species: Streptococcus dysgalactiae
C:Date: 28-Oct-1995 #sequence_revision 03-Nov-1995 #text_change 17-Mar-2000
C:Accession: S55890
R:Jonsson, H.; Frykberg, L.; Rantamaeki, L.; Guss, B.
Gene 143, 85-89, 1994
A:Title: MAG, a novel plasma protein receptor from Streptococcus dysgalactiae
A:Reference number: S55890; MUID:94259307
A:Accession: S55890
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-413 <ON>
A:Cross-references: EMBL:L27798; NID:g475117; PIDN:AAA26921.1; PID:g475118
C:Superfamily: M5 protein

```

Query Match	18.1%	Score	404;	DB	2;	Length	413;
Best Local Similarity	32.3%	Pred. No.	2.2e-14;				
Matches	141;	Conservative	31;	Mismatches	110;	Indels	154;
Gaps							
Qy	1	AVENKEETPETDSEEEV	TIKANLIFANGSTQTA	AEFKGTPEKATSEAYAYADTLKKDN	60		
Db	30	AVNAEESTVPTVATDAVT	-----	-----TSKEALATINKLISEDN	66		
Qy	61	GEYTVDVADKGYTLN	IKFAGKEKTPPEEKPEEVT	IKANL-IYADGKTQTA	AEFKGTPEEATA	119	
Db	67	LN-NLDI-----	-----QEVLLAKAGRDILASDSADT--	IKALLAEVTA	101		
Qy	120	EAYRYADALKKDN	GEYTVDVADKGYTLN	IKFAGKEKTPPEEKPEEVT	IKANLIYADGKTQT	179	
Db	102	EVTRL-----	NEEKARDAVDAKIAADAA-----	AFSELKD	132		
Qy	180	AEFKGTPEEATAEAYRYADL-	LAKENKTYTVDVADKGYTLN	IKFAGKEKTPPEEKPEEVT	1	238	
Db	133	AQLKAYEDLAKLAADTDL	DLWDVAKIINDYTRTKVENAKTAEDVK-----	KIFEESQNEVT-	186		
Qy	239	KANLIYADGKTQTA	AEFKGTFAEATAEAYRYADLLAKENKTYTADLEDGQYTNIRFAGKK	298			
Db	187	-----RIKTEKALKAALAKAKADAIE-	-----ILKKYG-----	IGDYI-	220		
Qy	299	VDEKPEEPMOTYKL	INGKTYLKGTEETTEAVDAATAEAKVPKQYANDNGVDGEWYTDATKT	358			
Db	221	-----KLINGKTAEGVT	-----	ALKD	237		

```

Qy 359 FTVTEKPEVIDASLSTPAVTTYKLVINGKTLKGTTTFAKVAETAFAKFOYANDNGVDG 418
      :  ||  |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
Db 238 EILASKPAVIDAPLTPALATYKLVINGKTLKGTTTFAKVAETAFAKFOYANENGVDG 297

Qy 419 VWTYDDATKTFVTTEM 434
      |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
Db 298 VWTYDDATKTFVTTEM 313

RESULT 8
A42404
collagen adhesin - Staphylococcus aureus
C:Species: Staphylococcus aureus
C:Date: 04-Mar-1993 #sequence_revision 18-Nov-1994 #text_change 10-Nov-1995
C:Accession: A42404; S27665
R:Patti, J.M.; Jonsson, H.; Guss, B.; Switalski, L.M.; Wiberg, K.; Lindberg, M.; Hook
J. Biol. Chem. 267, 4766-4772, 1992
A:Title: Molecular characterization and expression of a gene encoding a Staphylococcus
A:Reference number: A42404; MUID:92165839
A:Contents: FDA 574
A:Accession: A42404
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-1105 <PAT>
A:Cross-references: EMBL:M81736
A>Note: sequence extracted from NCBI backbone (NCBIP:83982)

```

Query Match	8.4%	Score 188.5	DB 2	Length 1185
Best Local Similarity	23.8%	Pred. No. 0.016		
Matches 136	Conservative 61	Mismatches 166	Indels 209	Gaps 38
Qy	2	VENKEETPETDSE-----EVTIKANLIIFANG-STOTAEFGTGFERAT 46		
Db	527	VTNTEKPIETTSISGEKWDKDNQDKRPEKSVN---LANGEKVYILD-----VT 576		
Qy	47	SEA---YAYADTLKKDNG-----EYTVDVADKGYTLNLIKFAKGEKTEPEEKE 90		
Db	577	SETNWKVEFKDLPKYDEGKIEYTVTEHDVXDYTTDI--NGTTIINKY-----TPGETSA 629		
Qy	91	EVT-----IKANLIYADGKTQ-----TAEFGTGFEEATAEAY 122		
Db	630	TVTKNWDDNNQDKRPEEIKVEL-YODGKATGTAILNESNNWTHTTGLDEKAKGOOV 688		
Qy	123	RY-ADALKKNGEYTVDVADKG-----YT-----LNIKFAKGEKTEPEEKEVTIKA 168		
Db	689	KYTVEELTKVGYYTHVDNNDMGNLIVTNKYTPETTSISGEKYWDKDNQDKRPEKYSV 748		
Qy	169	NLIYADG-KTQTAEFKGTFEERATA---YRYADLLAKENGK-----YTVDV 211		
Db	749	NLL-ADGEKVKRTL-----DVTSETNWKVEFKDLPKYDEGKIEYTVTEHDVXDYTTDI 800		
Qy	212	ADKGYTLNIKFAKGEKTEPEEKEEVT-----IKANLIYADGKTQ-----250		
Db	801	---NGTTIINKY-----TPGETSATVTKNWDNNQDKRPEEIKVEL-YODGKATGTAT 852		
Qy	251	-----TAEFGTFAEATAEAYRY-ADLLAKENGKYTDADLE---GGYVTNIRP-----294		
Db	853	LNESNNWTHTTGLDEKAKGOQVKYVEELTKVKG-YTHVDNNDMGNLIVTNKYTPETT 911		
Qy	295	--ACKKY-DEK-----PEPMDTYKILNGKTLKGTEETAEVDAATAEAKVFKQYANDN 344		
Db	912	SISGEKYWDKDNQDKRPEKV--SVNLLANGEKVK--TLDVTSETNWKVEFKDLPK--964		
Qy	345	GVDGWTVYDDATK-TFTVTEPEVIDASELTPAVTYTKLVINGKTLKGTEETTKAVDAETA 403		
Db	965	-----YDEGKKIEYTVTE-----DHVKDYTTDINGTTI-----TNKYTPGETS 1002		
Qy	404	EKAFKQYANDNGVDG-----VWYTDATKT 428		
Db	1003	ATVTKNWDNNQDKRPEEIKVELYODGKAT 1034		

RESULT 12

JC6171
late embryogenesis abundant protein - Arabidopsis thaliana
N;Alternate names: embryogenic cell protein 63; phosphotyrosine protein
C;Species: Arabidopsis thaliana (mouse-ear cress)
C;Date: 11-Apr-1997 #sequence_revision 09-May-1997 #text_change 21-Jul-2000
C;Accession: JC6171
R;Yang, H.; Saitou, T.; Kameda, Y.; Harada, H.; Kamada, H.
Gene 184, 83-88, 1997
A;Title: Arabidopsis thaliana ECP63 encoding a LEA protein is located in chromosome 4.
A;Reference number: JC6171; MUID: 97169149
A;Accession: JC6171
A;Molecule type: mRNA
A;Residues: 1-448 <YAN>
A;Cross-references: DDBJ:D64140; NID:g1526423; PIDN:BA011017.1; PID:g1526424
C;Comment: This protein is a phosphotyrosine protein which belongs to late embryogenesis
C;Genetics:
A;Gene: ATECP63
A;Map position: 4, south part
C;Keywords: seed

Query Match 7.5%; Score 168; DB 2; Length 448;
Best Local Similarity 24.2%; Pred. No. 0.059;
Matches 113; Conservative 50; Mismatches 161; Indels 142; Gaps 20;
QY 3 ENKEETPETDSEEVVTKANLIFANGSTQTAERFKGTPEKATSEAYAYADTLKKDNGE 62
DB 79 EAAESTKEGAIASEKAVGAK-----DATVEKAKET----ADYTAEKVGE 119
QY 63 YTVVDVADKGYTLNKFAGKEKTPPEKPEEVIKANIYADGKTQTAFKGTFFEAATAEAY 122
DB 120 Y-----KDYTVDAKAKAKOTTAETAKETANTYADKA-VEAKDKTAEKIGEYK----- 165
QY 123 RYADALKKNGEYTVVDVADKGYTLNKFAGKEKTPPEKPEEVIKANIYADGKTQTAF 182
DB 166 -----DYADKA-----VEAKDKTAETAKETANTYADKA-KEAKDKTAEK 204
QY 183 KGTPEETAEE-----AYRYADLLAKNGKYTVVDVADKGYTLNKFAGKEKTP 230
DB 205 VGEYKDYTVDAKAVDYEADYTAETAKAETAKETAGTGEY-----KDYTVDAKAKOTTA 258
QY 231 ---EPKEEV--TKANIYADGKTQTAFKGTFFEAATAEAYRYADLLAKNGKYTVADLE 285
DB 259 KLGLKDSAVETAKRAMGFLSGTTEAKGAVETKDT-----AKE----- 298
QY 286 GGYTINIRFAGKVDKPEPEPMDTYKLLNGKTLKGGTTTAAVDAAT-----AEKV 336
DB 299 -----NMEKAGEVTRQKME-----WRLEGKELKEAGAKAQAESQKTRESTESGAOKA 347
QY 337 --FKYANDNGVDGEYTVDDATKTFVTVEKPEVIDASEL-----TPAVTTYKL 382
DB 348 EETKDSAPVRGNEAKGTFGALGNVTETAIKSKLTMPDSIIVETRAAREHGSGTGVVEVK 407
QY 383 VINGKTLKGTTKYKAVDAETAEAKAFKQYANDNG--VDGVWYTVDDATK 427
DB 408 VEDSKPGKAVATSLKASQDMTG-----QTFNDVGRMD-----DDARK 443

RESULT 13
T52485
neurofilament protein NF-M(2), middle molecular weight [imported] - African clawed frog
C;Species: Xenopus laevis (African clawed frog)
C;Date: 24-Oct-2000 #sequence_revision 24-Oct-2000 #text_change 24-Oct-2000
C;Accession: T52485
R;Gervasi, C.; Szaro, B.G.
Brain Res. Mol. Brain Res. 48, 229-242, 1997
A;Title: Sequence and expression pattern of two forms of the middle molecular weight neu
A;Reference number: Z26090
A;Accession: T52485
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: mRNA
A;Residues: 1-913 <GER>
A;Cross-references: EMBL:085970; PIDN:AAB53390.1

Query Match 7.2%; Score 162; DB 2; Length 913;
Best Local Similarity 24.4%; Pred. No. 0.29;
Matches 119; Conservative 57; Mismatches 172; Indels 140; Gaps 23;
QY 3 ENKEETPETDSEEVVTKANLIFANGSTQTA-----EFKGTPEKATSEAY 50
DB 474 EEEEGEETQGEAEAEIVAAV-----ESSVQAAPGEEAEAEAEAEAEAEAE 530
QY 51 AYADTLKKNNGEYTVVDVADKGYTLNKFAGKEKTPPEK-----EVTIKANIYADGKT 105
DB 531 EKEEKEKEE-----DEG-----EKVDEEEGGEGEEDYKAE-----EEEK 571
QY 106 QTAEEKGTFFEEATAEAYRYADALKKNGEYTVVDVADKGYTLNKFAGKEKTPPEKPEEY 165
DB 572 EGEEDAEVEETKAE-----GEEGEGEDEAEK-----EGGEEEEEKEEVA 614
QY 166 IRANIYADGKTQTAFKGTFFEAATAEAYRYADLLAKE--NGKYTVVDVADKGYTLNKFAG 224
DB 615 -----AD-----EGGEGEEGEGEKEDEKGEKEGEAEAEAE-----GG 652
QY 225 KEKTPPEKPEEVIKANIYADGKTQTAFKGTFFA--EATAEAYRYADLLAKE--NGKY 279
DB 653 EETTEQAQAEVVEETMTETKIVREKAEVQTEAEAEETQDAEOKDEGEDEETGEGE 712
QY 280 TADLEGGYTIINIRFAG--KKVDEKPEEPMDTYKLLNGKTLKG----- 321
DB 713 TEKEEGEKEEDADSGTEQEEKAEKGE--IDEGETKEGAEDAGEDKQAKSP 769
QY 322 ETTTEAVDAATAEKYKQYANDNGVDGEYTVDDATKTFV-----TEKPEVIDASELT 377
DB 770 KSEETISYKSAASVSGSEKAKQEEKSTQRDCKGKEDTAVNGDVBEKGESEDKAVV 829
QY 378 TTYKLVING-----KTLKGETT---TKAVDAETAEKAFKQYANDNGVDGVWYTVDDA 425
DB 830 T-----NGLDSSPSDEKSSKGGQVTVTKKIEETTSE-----GEDGT---K 870
QY 426 TKTFVTVE 433
DB 871 TKSVTVTQ 878
RESULT 14
C95008
immunoglobulin A1 proteinase [imported] - Streptococcus pneumoniae (strain TIGR4)
C;Species: Streptococcus pneumoniae
C;Date: 03-Aug-2001 #sequence_revision 03-Aug-2001 #text_change 24-Aug-2001
C;Accession: C95008
R;Tetzelin, H.; Nelson, K.E.; Paulsen, I.T.; Eisen, J.A.; Read, T.D.; Peterson, S.; H
on, J.D.; Umayam, L.A.; White, O.; Salzberg, S.L.; Lewis, M.R.; Radune, D.; Holtzapp
nson, T.; Hickey, E.K.; Holt, I.E.
Science 293, 498-506, 2001
A;Authors: Loftus, B.J.; Yang, F.; Smith, H.O.; Venter, J.C.; Dougherty, B.A.; Morris
A;Title: Complete genome sequence of a virulent isolate of Streptococcus pneumoniae.
A;Reference number: A95000; MUID:21357209; PMID:11463916
A;Accession: C95008
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-1856 <KUR>
A;Cross-references: GB:AE005672; PIDN:AAK74260.1; PID:g14971537; GSPDB:GN00164; TIGR:
A;Experimental source: strain TIGR4
C;Genetics:
A;Gene: SP0071
C;Superfamily: Streptococcus sanguis IgA-specific metalloendopeptidase
Query Match 7.1%; Score 159; DB 2; Length 1856;
Best Local Similarity 23.1%; Pred. No. 1;
Matches 119; Conservative 60; Mismatches 176; Indels 160; Gaps 28;
QY 7 ETPETPETDSEEVVTKANLIFANGSTQTAERFKGT-----FEKAT-SEAYAYAD---- 54
DB 315 ETGTOEPCHGEAAVREDLPVY-----TKPLETKGTGPGHGEAAVREEEPAYTEPLAT 370
QY 55 -----TLKKDNGEYTVVDVADKGYTLNKFAGKEKTPPE--PKEEVTK-----A 96

Db 371 KGTQEPGHEGKATVREETLEYTPVATKG-TQBPHEGEAAVEELPALEVTTRNRTEIQ 429
QY 97 NLIYADGKTQ-----TAERKGFEEATAEAYRYADALKKDNCEYT-----VDYADK 142
Db 430 NIPYTEEIODPTLLKNRRKIEROG---QAGTTIQIEDYIVNGNVVETKEVSRTEVAPV 486
QY 143 GYTLNKFAGKEK-----TPEEPKEEVTIRKANLI-----YADGKTQTAERKGFTEEA 189
Db 487 NEVVKVGTLVKVPTVEITNLTKVENKKSITVSNLIDTTSAIVSAKTQV--FHG----- 539
QY 190 TAEAYRYADLLAKENGYTVDVADKGYTLNKFAGKEKTPPEEP-KEEV-----TI 238
Db 540 -----DKLVKE-----VDI-----ENPAKEQVISGLDYTYPTV 568
QY 239 KANLIYADG-----KTQTAERKGFEEATAEAYRYADLLAKENGYTADLEDGGTYIN 291
Db 569 KTHLTYNLNGNNEENTSTQDQLEYKKEIKDIDSVELYKENDRY----- 616
QY 292 IRFAGKVKDEKPEPMDTYKLIILNGKT-----LKGETTTEAVDAATAEKVFKQYANDN 344
Db 617 -----RRYLSLSEAPTDTAKYFVKVKSDEKEMVLPVKSITENTDGYKYVTVAVDQVLEE 671
QY 345 GVDGEYTYDATKTFV-----TEKPEVIDASELTPAVTYTKLVING-KTLKGETTTTKAYD 399
Db 672 GTDG---YKD-DYTFVAKSKAEPQGVY--TSFKQLVTAMQSNLSGVYTLASDWTADDEV 725
QY 400 AETAERAKFKQYANDNGVDGVWYVDDATKTFVTTEM 434
Db 726 LGDKQSYLTGAFGTGSLIG-----SDGPKSAIYDL 756

RESULT 15

B98047

hypothetical protein spr1403 [imported] - Streptococcus pneumoniae (strain R6)

C:Species: Streptococcus pneumoniae

C:Date: 22-Oct-2001 #sequence_revision 22-Oct-2001 #text_change 22-Oct-2001

C:Accession: B98047

R:Hoskins, J.A.; Alborn Jr., W.; Arnold, J.; Blaszcak, L.; Burgett, S.; DeHoff, B.S.; E

e, R.; LeBlanc, D.J.; Lee, L.N.; Lefkowitz, E.J.; Lu, J.; Matsushima, P.; McAhren, S.; M

y, P.; Sun, P.M.; Winkler, M.E.

J. Bacteriol. 183, 5709-5717, 2001

A:Authors: Yang, Y.; Young-Bellido, M.; Zhao, G.; Zook, C.; Baltz, R.H.; Jaskunas, S.R.;

A>Title: Genome of the Bacterium Streptococcus pneumoniae Strain R6.

A:Reference number: A97872; MUID:21429245; PMID:11544234

A:Accession: B98047

A>Status: preliminary

A:Molecule type: DNA

A:Residues: 1-2551 <KUR>

A:Cross-references: GB:AE007317; PIDN:AAL00207.1; PID:g15459054; GSPDB:GN00174

C:Genetics:

A:Gene: spr1403

Query Match

Best Local Similarity 7.1%; Score 159; DB 2; Length 2551;

Matches 107; Conservative 46; Mismatches 162; Indels 150; Gaps 21;

QY 2 VENKEETPETPDSEEEVTIKANLIFANGSTQTAERKGFTEKATSEAYAYADTLKKDNG 61
Db 1965 IDRGDSAPTITTKDNGDGHHTITVQNPDGSESTVYVKGDKGKT----ANITTTENPDG 2020
QY 62 EYTVDVADKGYTLNKFAGKEKTPPEEKV-----TIK-----ANLIYADGKT 105
Db 2021 SHITVTPDGSRKETVVKNGKDGKTKPEVETDNDGTHTVKVDGNGVNTNALIKDKD 2080
QY 106 QTAERKGFTEATAEAYRYADALKKONGEYTVDVADKGYTLNKFAGKEKTPPEEKV 165
Db 2081 GKA-----ATATTTENPD-----GSHTVTITN-----PDGKNKEFV 2111
QY 166 IKANLIYADGKTQTAERKGFTEATAEAYRYADLLAKENGYTVDVADKGYTLNKFAGK 225
Db 2112 VKNGRDVGDRTPTAS-----VRDNG-----DGSHTITVI----- 2140

QY 226 EKTPPEEPKEEVTIRKANLIYADGK-----TQTAERKGFTEATAEAYRYADLLAKENGK--- 278
Db 2141 -TNPEGVTTTETTVR-----DGKSPKVTITDEQNGTHKISVLNGDGTTFETIIKDGKSPV 2193
QY 279 -YTADLEDGCGYTTNIRFAGKKVDEKPEPMDTYKLIILNGKTLKGETTTEAVDAATAEKVF 337
Db 2194 ATVRDNDQDGYTTIRVE-----NNGTIVSETTVRDGKSPTAKVV- 2231
QY 338 KQYANDNGVDGEWY-----DDATKTFVTVE-----KPEVIDASELTPAVTYTKLVINGKT 388
Db 2232 -----DMG-DGTHITVTVVNSDGIITTTTVRDGREGPKLEVIDNND-----GSHTIKVTVGAD 2280
QY 389 LKGETTTTKAYDAETAERKAFKQYANDNGVDGVWYVDDATKTFVTVE 433
Db 2281 GKG-TTTTIFDG-----KSPKANIVDNG-----DGTHLTITVD 2312

Search completed: October 29, 2002, 09:31:33

Job time : 24.2057 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.3
Copyright (c) 1993 - 2002 Compugen Ltd.

OM protein - protein search, using sw model

Run on: October 29, 2002, 09:28:27 : Search time 13.5074 Seconds
(without alignments)
784.805 Million cell updates/sec

Title: US-08-325-278B-3

Perfect score: 2235

Sequence: 1 AVENKEETPETPTDSEEV.....GVDGVWVYDDATKFTVTM 434

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 231628 seqs, 24425594 residues

Total number of hits satisfying chosen parameters: 231628

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Issued_Patents_AA.*

- 1: /cgn2_6/ptodata/1/iaa/5A_COMB.pep.*
- 2: /cgn2_6/ptodata/1/iaa/5B_COMB.pep.*
- 3: /cgn2_6/ptodata/1/iaa/6A_COMB.pep.*
- 4: /cgn2_6/ptodata/1/iaa/6B_COMB.pep.*
- 5: /cgn2_6/ptodata/1/iaa/PCTUS_COMB.pep.*
- 6: /cgn2_6/ptodata/1/iaa/backfiles1.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	ID	Description
1	2235	100.0	434	2	US-08-795-475-3
2	1565	70.0	305	2	US-08-795-475-1
3	1263.5	56.5	1027	4	US-08-446-137B-2
4	1216	54.4	291	4	US-08-446-137B-4
5	706.5	31.6	664	3	US-08-669-408B-2
6	622.5	27.9	502	1	US-08-378-761A-25
7	622.5	27.9	502	1	US-08-485-286-25
8	622	27.8	493	1	US-08-378-761A-23
9	622	27.8	493	1	US-08-485-286-23
10	618	27.7	342	3	US-08-828-741B-6
11	618	27.7	342	4	US-09-160-567-6
12	618	27.7	489	1	US-08-378-761A-19
13	618	27.7	489	1	US-08-485-286-19
14	615	27.5	178	3	US-08-828-741B-13
15	615	27.5	178	4	US-09-160-567-13
16	615	27.5	198	3	US-08-828-741B-8
17	615	27.5	198	4	US-09-160-567-8
18	614	27.5	495	3	US-08-828-741B-4
19	614	27.5	495	4	US-09-160-567-4
20	613	27.4	182	3	US-08-828-741B-2
21	613	27.4	182	4	US-09-160-567-2
22	404	18.1	413	3	US-08-669-408B-10
23	369.5	16.5	402	1	US-08-378-761A-17
24	369.5	16.5	402	1	US-08-485-286-17
25	334	14.9	75	4	US-08-446-137B-8
26	315.5	14.1	71	4	US-08-446-137B-6
27	306	13.7	74	4	US-08-446-137B-7

28	303	13.6	71	4	US-08-446-137B-5	Sequence 5, Appli
29	293	13.1	57	4	US-09-117-233-16	Sequence 16, Appli
30	292	13.1	57	4	US-09-117-233-2	Sequence 2, Appli
31	292	13.1	60	4	US-09-117-233-12	Sequence 12, Appli
32	287	12.8	56	4	US-09-058-459-38	Sequence 38, Appli
33	287	12.8	56	4	US-09-127-926-38	Sequence 8, Appli
34	286	12.8	57	4	US-09-117-233-8	Sequence 10, Appli
35	286	12.8	58	4	US-09-117-233-10	Sequence 1, Appli
36	286	12.8	58	4	US-09-444-305A-1	Sequence 6, Appli
37	286	12.8	60	4	US-09-117-233-6	Sequence 4, Appli
38	285	12.8	60	4	US-09-117-233-4	Sequence 68, Appli
39	283	12.7	56	4	US-09-058-459-68	Sequence 68, Appli
40	283	12.7	56	4	US-09-127-926-68	Sequence 43, Appli
41	281	12.6	56	4	US-09-058-459-46	Sequence 46, Appli
42	280	12.5	56	4	US-09-058-459-43	Sequence 43, Appli
43	280	12.5	56	4	US-09-058-459-44	Sequence 44, Appli
44	280	12.5	56	4	US-09-058-459-58	Sequence 58, Appli
45	280	12.5	56	4	US-09-058-459-70	Sequence 70, Appli

ALIGNMENTS

RESULT 1
US-08-795-475-3
: Sequence 3, Application US/08795475
: Patent No. 5965390
: GENERAL INFORMATION:
: APPLICANT: Bjvick, Lars
: APPLICANT: Sjvbring, Ulf
: TITLE OF INVENTION: PROTEIN L AND HYBRID PROTEINS THEREOF
: NUMBER OF SEQUENCES: 14
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: SEED AND BERRY LLP
: STREET: 6300 Columbia Center, 701 Fifth Avenue
: CITY: Seattle
: STATE: Washington
: COUNTRY: USA
: ZIP: 98104-7092
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: COMPUTER: IBM PC compatible
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: PatentIn Release #1.0, Version #1.30
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/795,475
: FILING DATE: 11-FEB-1997
: CLASSIFICATION: 435
: ATTORNEY/AGENT INFORMATION:
: NAME: Mcmasters, David D.
: REGISTRATION NUMBER: 33,963
: REFERENCE/DOCKET NUMBER: 100084.402D1
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: (206) 622-4900
: TELEFAX: (206) 682-6031
: INFORMATION FOR SEQ ID NO: 3:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 434 amino acids
: TYPE: amino acid
: STRANDEDNESS: unknown
: TOPOLOGY: unknown
: MOLECULE TYPE: protein
: HYPOTHETICAL: NO
: ORIGINAL SOURCE:
: ORGANISM: Escherichia coli LE392/pHDLG, DSM 7055
US-08-795-475-3

2
DMS
+
calc
*
2
p
p
p

Query Match 100.0%; Score 2235; DB 2; Length 434;
Best Local Similarity 100.0%; Pred. No. 1.8e-170; Indels 0; Gaps 0;
Matches 434; Conservative 0; Mismatches 0;
QY 1 AVENKEETPETPTDSEEVTKANLIPANGSTQAEKFTFKATSEAYADTLKKDN 60
|||||

Db 1 AVENKEETPETDSEBEVITKANLIFANGSTQTAEFKGTFFKATSEAYAYADTLKKN 60
Qy 61 GEYTVADVADKGYTLNIFAGKEKTPPEPKKEVITKANLIYADGKTQTAEFKGTFFKATSEAYAYADTLKKN 120
Db 61 GEYTVADVADKGYTLNIFAGKEKTPPEPKKEVITKANLIYADGKTQTAEFKGTFFKATSEAYAYADTLKKN 120
Qy 121 AYRYADALKKONGEYTVADVADKGYTLNIFAGKEKTPPEPKKEVITKANLIYADGKTQTA 180
Db 121 AYRYADALKKONGEYTVADVADKGYTLNIFAGKEKTPPEPKKEVITKANLIYADGKTQTA 180
Qy 181 EFKGTFFKATSEAYAYADTLKKNKGYTVADVADKGYTLNIFAGKEKTPPEPKKEVITKANLIYADGKTQTA 240
Db 181 EFKGTFFKATSEAYAYADTLKKNKGYTVADVADKGYTLNIFAGKEKTPPEPKKEVITKANLIYADGKTQTA 240
Qy 241 NLIYADGKTQTAEFKGTFFKATSEAYAYADTLKKNKGYTVADVADKGYTLNIFAGKEKTPPEPKKEVITKANLIYADGKTQTA 300
Db 241 NLIYADGKTQTAEFKGTFFKATSEAYAYADTLKKNKGYTVADVADKGYTLNIFAGKEKTPPEPKKEVITKANLIYADGKTQTA 300
Qy 301 EKPEPMDTYKLIILNGKTLKGETTTEAVDAATAEKVFKQYANDNGVDGWTYDDATKTFT 360
Db 301 EKPEPMDTYKLIILNGKTLKGETTTEAVDAATAEKVFKQYANDNGVDGWTYDDATKTFT 360
Qy 361 VTERPEVIDASELPVATYTKLVINGKTLKGETTTEAVDAATAEKVFKQYANDNGVDGWTYDDATKTFT 420
Db 361 VTERPEVIDASELPVATYTKLVINGKTLKGETTTEAVDAATAEKVFKQYANDNGVDGWTYDDATKTFT 420
Qy 421 TYDDATKTFTVTEM 434
Db 421 TYDDATKTFTVTEM 434

RESULT 2
US-08-795-475-1
; Sequence 1, Application US/08795475
; Patent No. 5965390
; GENERAL INFORMATION:
; APPLICANT: Bjvick, Lars
; APPLICANT: Sjbvring, Ulf
; TITLE OF INVENTION: PROTEIN L AND HYBRID PROTEINS THEREOF
; NUMBER OF SEQUENCES: 14
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: SEED and BERRY LLP
; STREET: 6300 Columbia Center, 701 Fifth Avenue
; CITY: Seattle
; STATE: Washington
; COUNTRY: USA
; ZIP: 98104-7092
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/795,475
; FILING DATE: 11-FEB-1997
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: McMasters, David D.
; REGISTRATION NUMBER: 33,963
; REFERENCE/DOCKET NUMBER: 100084.402D1
; TELEPHONE: (206) 622-4900
; TELEFAX: (206) 682-6031
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 305 amino acids
; TYPE: amino acid
; STRANDEDNESS: unknown
; TOPOLOGY: unknown
; MOLECULE TYPE: protein
; HYPOTHETICAL: NO
; ORIGINAL SOURCE:
; ORGANISM: Escherichia coli LE392/pHDL, DSM 7054

US-08-795-475-1

Query Match 70.0%; Score 1565; DB 2; Length 305;
Best Local Similarity 100.0%; Pred. No. 2,6e-117;
Matches 305; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 AVENKEETPETDSEBEVITKANLIFANGSTQTAEFKGTFFKATSEAYAYADTLKKN 60
Db 1 AVENKEETPETDSEBEVITKANLIFANGSTQTAEFKGTFFKATSEAYAYADTLKKN 60
Qy 61 GEYTVADVADKGYTLNIFAGKEKTPPEPKKEVITKANLIYADGKTQTAEFKGTFFKATSEAYAYADTLKKN 120
Db 61 GEYTVADVADKGYTLNIFAGKEKTPPEPKKEVITKANLIYADGKTQTAEFKGTFFKATSEAYAYADTLKKN 120
Qy 121 AYRYADALKKONGEYTVADVADKGYTLNIFAGKEKTPPEPKKEVITKANLIYADGKTQTA 180
Db 121 AYRYADALKKONGEYTVADVADKGYTLNIFAGKEKTPPEPKKEVITKANLIYADGKTQTA 180
Qy 181 EFKGTFFKATSEAYAYADTLKKNKGYTVADVADKGYTLNIFAGKEKTPPEPKKEVITKANLIYADGKTQTA 240
Db 181 EFKGTFFKATSEAYAYADTLKKNKGYTVADVADKGYTLNIFAGKEKTPPEPKKEVITKANLIYADGKTQTA 240
Qy 241 NLIYADGKTQTAEFKGTFFKATSEAYAYADTLKKNKGYTVADVADKGYTLNIFAGKEKTPPEPKKEVITKANLIYADGKTQTA 300
Db 241 NLIYADGKTQTAEFKGTFFKATSEAYAYADTLKKNKGYTVADVADKGYTLNIFAGKEKTPPEPKKEVITKANLIYADGKTQTA 300
Qy 301 EKPEE 305
Db 301 EKPEE 305

RESULT 3
US-08-446-137B-2
; Sequence 2, Application US/08446137B
; Patent No. 6162903
; GENERAL INFORMATION:
; APPLICANT: Trowern, Angus R.
; APPLICANT: Atkinson, Anthony
; APPLICANT: Murphy, Jonathan P.
; APPLICANT: Laurence, Oliver S.
; APPLICANT: Duggleby, Clive J.
; TITLE OF INVENTION: IMMUNOGLOBULIN BINDING PROTEINS DERIVED
; TITLE OF INVENTION: FROM L PROTEIN AND THEIR USES
; NUMBER OF SEQUENCES: 12
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: SEED and BERRY LLP
; STREET: 6300 Columbia Center, 701 Fifth Avenue
; CITY: Seattle
; STATE: Washington
; COUNTRY: USA
; ZIP: 98104
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/446,137B
; FILING DATE: 22-MAY-1995
; CLASSIFICATION: 514
; ATTORNEY/AGENT INFORMATION:
; NAME: McMasters, David D.
; REGISTRATION NUMBER: 33,963
; REFERENCE/DOCKET NUMBER: 100084.406
; TELEPHONE: (206) 622-4900
; TELEFAX: (206) 682-6031
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1027 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein


```

; ; REFERENCE/DOCKET NUMBER: 61743/102
; ;
; ; TELECOMMUNICATION INFORMATION:
; ;
; ; TELEPHONE: (202)672-5300
; ; TELEFAX: (202)672-5399
; ; TELEX: 904136
; ;
; ; INFORMATION FOR SEQ ID NO: 2:
; ;
; ; SEQUENCE CHARACTERISTICS:
; ;
; ; LENGTH: 664 amino acids
; ; TYPE: amino acid
; ; TOPOLOGY: linear
; ; MOLECULE TYPE: protein
; ;
; ; US-08-669-408B-2

```

Query Match 31.6%; Score 706.5; DB 3; Length 664;
Best Local Similarity 43.9%; Pred. NO. 1.9e-48;
Matches 192; Conservative 51; Mismatches 143; Indels 51; Gaps 17;

RESULT 6
US-08-378-761A-25
; Sequence 25, Application US/08378761A

```

; SOFTWARE: PatentIn Release #1.0, Version #1.25
;
; CURRENT APPLICATION DATA:
;
; APPLICATION NUMBER: US/08/378,761A
; FILING DATE: 26-JAN-1995
; CLASSIFICATION: 435
;
; ATTORNEY/AGENT INFORMATION:
;
; NAME: BORUCKI, ANDREA T
; REGISTRATION NUMBER: 33651
; REFERENCE/DOCKET NUMBER: 38272B
;
; TELECOMMUNICATION INFORMATION:
;
; TELEPHONE: (317) 337-4846
;
; INFORMATION FOR SEQ ID NO: 25:
;
; SEQUENCE CHARACTERISTICS:
; LENGTH: 502 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
;
; US-08-378-761A-25

```

Query Match 27.9%; Score 622.5; DB 1; Length 502;
Best Local Similarity 41.1%; pred. No. 6.3e-42;
Matches 178; Conservative 36; Mismatches 136; Indels 83; Gaps 16;

```
Qy 37 EFKGF-----EKATSEAYADTLKNDGEYTDV-ADKGYTLNKFAGKEKTPPEPK 89
      :|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||
Db 38 DHKGFQVLPPEKKVPPELWPFYTB-LKTRTSITLARMONLYLUGFRTPG-----87
      :|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||
Qy 90 EEVTIKANLIYADGKTQTAFFKGTFEEATAEAYADALKDKNDGEYTDVADKGYTLNIK 149
      :|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||
Db 88 -----GVWWEFGCKDGTLLGDNPRWLGGGGRYQD-LIGNKGLGVTVMGRAEMTRAVN 139
      :|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||
Qy 150 FAGKEKTPPEPKBEVTIKANLIYADGKTQTAFFKGTFEEATAEAYRYADLLAKENGKTYV 209
      :|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||
Db 140 DLAKKKKAADPQ-----ADTKSLVK----LVVMVCEGLRNTV-----SRTV 178
      :|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||
Qy 210 DV----ADKGYTLNKFAGKEKTPPEKPEEVTIKANLIYADGKTQT-AEFKGTFPAETAEEA 265
      :|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||
Db 179 DAGNSOHGVTLVTV-OGK----OVQKWDRIKKAAPFWDHPTAVIPDMQKLGTKDKNEA 233
      :|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||
Qy 266 YRYADLLAKENGKYTADLEGGYTVINIRFAGKVKDEKPE-----EPMDTYKLIINGK 317
      :|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||
Db 234 ARIVALV--KNQTTAAATAAGSSCARVRRSCGYD-RPEVIDASELTPAVTYYKLVINGK 290
      :|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||
Qy 318 TLKGETTTEAVDAATAEKVKQYANDNGVGEWYDDATKTFVTYE-----363
      :|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||
Db 291 TLKGETTTEAVDAATAEKVKQYANDNGVGEWYDDATKTFVTTEKPEVIDASELTPAV 350
      :|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||
Qy 364 ---KPEVIDASELTPAVTYYKLVINGKTLKGETTTKAVDAETAEKVKQYANDNGVDGVW 420
      :|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||
Db 351 TRSPKPEVIDASELTPAVTYYKLVINGKTLKGETTTEAVDAATAEKVKQYANDNGVDGEW 410
      :|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||
```

RESULT 7
US-08-485-286-25
; Sequence 25, Application US/08485286
; Patent No. 5646026
; Patent No. 5646026 5646119
; GENERAL INFORMATION:
; APPLICANT: WALSH, TERENCE A
; APPLICANT: HEY, TIMOTHY D
; APPLICANT: MORGAN, ALICE ER
; TITLE OF INVENTION: RIBOSOME-INACT
; TITLE OF INVENTION: PRECURSOR FOR
; TITLE OF INVENTION: USING
; NUMBER OF SEQUENCES: 81
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: ANDREA T. BORUCKI
; STREET: 9330 ZIONSVILLE ROAD
; CITY: INDIANAPOLIS

; STATE: IN
 ; COUNTRY: US
 ; ZIP: 46268
 ; COMPUTER READABLE FORM:
 ; MEDIUM TYPE: Floppy disk
 ; COMPUTER: IBM PC compatible
 ; OPERATING SYSTEM: PC-DOS/MS-DOS
 ; SOFTWARE: Patent In Release #1.0, Version #1.25
 ; CURRENT APPLICATION DATA:
 ; APPLICATION NUMBER: US/08/485,286
 ; FILING DATE:
 ; CLASSIFICATION: 435
 ; PRIOR APPLICATION DATA:
 ; APPLICATION NUMBER: US 08/378761
 ; FILING DATE: 26-JAN-1995
 ; ATTORNEY/AGENT INFORMATION:
 ; NAME: BORUCKI, ANDREA T
 ; REGISTRATION NUMBER: 33651
 ; REFERENCE/DOCKET NUMBER: 38272B
 ; TELECOMMUNICATION INFORMATION:
 ; TELEPHONE: (317) 337-4846
 ; INFORMATION FOR SEQ ID NO: 25:
 ; SEQUENCE CHARACTERISTICS:
 ; LENGTH: 502 amino acids
 ; TYPE: amino acid
 ; TOPOLOGY: linear
 ; MOLECULE TYPE: protein
 ; US-08-485-286-25

Query Match 27.9%; Score 622.5; DB 1; Length 502;
 Best Local Similarity 41.1%; Pred. No. 6.3e-42;
 Matches 178; Conservative 36; Mismatches 136; Indels 83; Gaps 16;

QY	37	EFGKTF-----EKATSEAYAYADTLKKDNGEYTVDVV-ADKGYTLNKKFAGKEKTPPEPK	89
DB	38	DHKGIFQVLPPEKVPPELWFYTE-LKTRTSSITLAIMDNLYLVGPRTEG-----	87
QY	90	EEVTIKANLIYADGKTQTAEFGTFEEATAEAYRYADALKKNGEYTVDVVADKGYTLNKK	149
DB	88	-----GVWVEFGKGDTHLLGNPWLFGGGRYQD-LIGNKGLVTVWGRAEMTRAVN	139
QY	150	FAGKEKTPPEKPEVTIKANLIYADGKTQTAEFGTFEEATAEAYRYADALKKNGEYTV	209
DB	140	DLAKKKKAADPQ-----ADTKSLVK-----LVWVCEGLRNTV-----SRTV	178
QY	210	DV----ADKGYTLNKKFAGKEKTPPEKPEVTIKANLIYADGKTQT-AEFGKTFEATAEAB	265
DB	179	DAGFNSQGVTLTVT-QGK-----QVQKWDRIKAAFEAWDHPTAVIPDMQKLGKDKNEA	233
QY	266	RYADLLAKENGKYTADLEDGGYTTINIRFAGKKVDEKPE-----EPMDTYKLIINGK	317
DB	234	ARIVALV--KNQTTAAATAGSSCARVRSRSCGVV--KPEVIDASELTPAVTTYKLVI	290
QY	318	TLKGETTTEAVDAATAEAKVFKQYANDNGVDGEWYDDATKTFVTVE-----	363
DB	291	TLKGETTTEAVDAATAEAKVFKQYANDNGVDGEWYDDATKTFVTTEKPEVIDASELTPAV	350
QY	364	---KPEVIDASELTPAVTTYKLVIINGKTLKGETTTKAVDAEATAEAKVFKQYANDNGV	420
DB	351	TRSKPEVIDASELTPAVTTYKLVIINGKTLKGETTTEAVDAATAEAKVFKQYANDNGV	410
QY	421	TYDDATKTFVTVE 433	
DB	411	TYDDATKTFVTVE 423	

RESULT 8
 US-08-378-761A-23
 ; Sequence 23, Application US/08378761A
 ; Patent No. 5635384
 ; GENERAL INFORMATION:
 ; APPLICANT: WALSH, TERENCE A
 ; APPLICANT: HEY, TIMOTHY D

; APPLICANT: MORGAN, ALICE ER
 ; TITLE OF INVENTION: RIBOSOME-INACTIVATING PROTEINS, INACTIVE
 ; TITLE OF INVENTION: PRECURSOR FORMS THEREOF, A PROCESS FOR MAKING A METHOD OF
 ; NUMBER OF SEQUENCES: 81
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: ANDREA T. BORUCKI
 ; STREET: 9330 ZIONSVILLE ROAD
 ; CITY: INDIANAPOLIS
 ; STATE: IN
 ; COUNTRY: US
 ; ZIP: 46268
 ; COMPUTER READABLE FORM:
 ; MEDIUM TYPE: Floppy disk
 ; COMPUTER: IBM PC compatible
 ; OPERATING SYSTEM: PC-DOS/MS-DOS
 ; SOFTWARE: Patent In Release #1.0, Version #1.25
 ; CURRENT APPLICATION DATA:
 ; APPLICATION NUMBER: US/08/378,761A
 ; FILING DATE: 26-JAN-1995
 ; CLASSIFICATION: 435
 ; ATTORNEY/AGENT INFORMATION:
 ; NAME: BORUCKI, ANDREA T
 ; REGISTRATION NUMBER: 33651
 ; REFERENCE/DOCKET NUMBER: 38272B
 ; TELECOMMUNICATION INFORMATION:
 ; TELEPHONE: (317) 337-4846
 ; INFORMATION FOR SEQ ID NO: 23:
 ; SEQUENCE CHARACTERISTICS:
 ; LENGTH: 493 amino acids
 ; TYPE: amino acid
 ; TOPOLOGY: linear
 ; MOLECULE TYPE: protein
 ; US-08-378-761A-23

Query Match 27.8%; Score 622; DB 1; Length 493;
 Best Local Similarity 46.9%; Pred. No. 6.7e-42;
 Matches 146; Conservative 28; Mismatches 55; Indels 82; Gaps 9;

QY	195	RYADLLAKENGKYTVDVADKGYTLNKKFAGKEKTPPEKPEV---TIKANLIYADG----	247
DB	114	RYQDLGNK-GLETVTWGRAEMTRAVNDLAKKKADPOADTKSLKLVVMVCEGLRPN	172
QY	248	---KTQTAEPFG-----TFEATAEAYRYAD-----	270
DB	173	TVSTVDAGFNSQGVTLTVTQGVQVQKWDRIKAAFEAWDHPTAVIPDMQKLGKDKNE	232
QY	271	---LLAKENGKYTADLEDGGYTTINIRFAGKKVDEKPE-----EPMDTYKLIINGKTL	319
DB	233	AARIVALVKNQTTAAATAG-SVNV-----DKPEVIDASELTPAVTTYKLVI	283
QY	320	KGETTTEAVDAATAEAKVFKQYANDNGVDGEWYDDATKTFVTVE-----	363
DB	284	KGETTTEAVDAATAEAKVFKQYANDNGVDGEWYDDATKTFVTTEKPEVIDASELTPAVTR	343
QY	364	-KPEVIDASELTPAVTTYKLVIINGKTLKGETTTKAVDAEATAEAKVFKQYANDNGV	422
DB	344	SKPEVIDASELTPAVTTYKLVIINGKTLKGETTTEAVDAATAEAKVFKQYANDNGV	403
QY	423	DDATKTFVTVE 433	
DB	404	DDATKTFVTVE 414	

RESULT 9
 US-08-485-286-23
 ; Sequence 23, Application US/08485286
 ; Patent No. 5646026
 ; Patent No. 5646026 5646119
 ; GENERAL INFORMATION:
 ; APPLICANT: WALSH, TERENCE A
 ; APPLICANT: HEY, TIMOTHY D
 ; APPLICANT: MORGAN, ALICE ER

```

; TITLE OF INVENTION: RIBOSOME-INACTIVATING PROTEINS, INACTIVE
; TITLE OF INVENTION: PRECURSOR FORMS THEREOF, A PROCESS FOR MAKING A METHOD OF
; TITLE OF INVENTION: USING
; NUMBER OF SEQUENCES: 81
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: ANDREA T. BORUCKI
; STREET: 9330 ZIONSVILLE ROAD
; CITY: INDIANAPOLIS
; STATE: IN
; COUNTRY: US
; ZIP: 46268
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/485,286
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/378761
; FILING DATE: 26-JAN-1995
; ATTORNEY/AGENT INFORMATION:
; NAME: BORUCKI, ANDREA T
; REGISTRATION NUMBER: 33651
; REFERENCE/DOCKET NUMBER: 38272B
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (317) 337-4846
; INFORMATION FOR SEQ ID NO: 23:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 493 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-485-286-23

```

```

Query Match 27.8%; Score 622; DB 1; Length 493;
Best Local Similarity 46.9%; Pred. No. 6.7e-42;
Matches 146; Conservative 28; Mismatches 55; Indels 82; Gaps 9;

QY 195 RYADLLAKENGKTYVDVADKGYTLNFKAGKEKTPPEKPEV---TKANLIYADG---- 247
Db 114 RYQDLINCK-GLETVTMGRAEMTRAVNDLAKKKAADPQADTKSKLVKLVVMVCEGLRFN 172

QY 248 ---KTQTAEFKG-----TFEAATAEAYRYAD----- 270
Db 173 TVSRVTDAAGFNSQHGVTTLVTQCKQVQKQWDRISKAAFEWADHPTAVIPDMQKLGIDKNE 232

QY 271 ---LLAKENGKYTADLEDGGYTTINIRPAGKKVDEKPE-----EPMDTYKILANGKTL 319
Db 233 AARIVALVKQNTTAAATAAG-SVNV-----DKPEVIDASELTPAVTYTKLVINGKTL 283

QY 320 KGETTTEAVDAATAEKVKFYQYANDNGVDGWTYDDATKFTVTVE----- 363
Db 284 KGETTTEAVDAATAEKVKFYQYANDNGVDGWTYDDATKFTVTVEKPEVIDASELTPAVTR 343

QY 364 -KPEVIDASELTPAVTYTKLVINGKTLKGETTKAVDAETAERKAFKQYANDNGVDGWTY 422
Db 344 SKPEVIDASELTPAVTYTKLVINGKTLKGETTEAVDAATAEKVKFYQYANDNGVDGWTY 403

QY 423 DDAKTKFTVTE 433
Db 404 DDAKTKFTVTE 414

RESULT 10
US-08-828-741B-6
; Sequence 6, Application US/08828741B
; Patent No. 6043069
; GENERAL INFORMATION:
; APPLICANT: Koentgen, Frank
; APPLICANT: Suess, Gabriele M.

```

```

; APPLICANT: Tarlinton, David M.
; APPLICANT: Treutlein, Herbert R.
; TITLE OF INVENTION: CATALYTIC ANTIBODIES AND A METHOD OF
; TITLE OF INVENTION: PRODUCING SAME
; NUMBER OF SEQUENCES: 14
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: SCULLY, SCOTT, MURPHY & PRESSER
; STREET: 400 Garden City Plaza
; CITY: Garden City
; STATE: New York
; COUNTRY: United States of America
; ZIP: 11530
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/828,741B
; FILING DATE: 26-MAR-1997
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Digiglio, Frank S.
; REGISTRATION NUMBER: 31,346
; REFERENCE/DOCKET NUMBER: 10591
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (516) 742-4343
; TELEFAX: (516) 742-4366
; TELEX: 230 901 SANS UR
; INFORMATION FOR SEQ ID NO: 6:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 342 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-828-741B-6

Query Match 27.7%; Score 618; DB 3; Length 342;
Best Local Similarity 75.9%; Pred. No. 8.4e-42;
Matches 126; Conservative 5; Mismatches 15; Indels 20; Gaps 1;

QY 16 SEEVTVKANLIPANGSTQTAEPKGFTEKATSEAYAYADTLKKDNGEYTVDVADKGYTLN 75
Db 174 SAEVTVKANLIPANGSTQTAEPKGFTEKATSEAYAYADTLKKDNGEYTVDVADKGYTLN 233

QY 76 IKFAGKEKTPPE-----PKEEVTIKANLIYADGKTQTAEFKGTPE 115
Db 234 IKFAGKEATNRNTDGSYDYLQINSRWGLTSABEVTIKANLIFANGSTQTAEPKGTPE 293

QY 116 EATAEAYRYADALKKNGEYTVDVADKGYTLNFKAGKEKTPPEPK 161
Db 294 KATSEAYAYADTLKKDNGEYTVDVADKGYTLNFKAGKESAWRHPQ 339

RESULT 11
US-09-160-567-6
; Sequence 6, Application US/09160567
; Patent No. 6326179
; GENERAL INFORMATION:
; APPLICANT: Koentgen, Frank
; APPLICANT: Suess, Gabriele M.
; APPLICANT: Tarlinton, David M.
; APPLICANT: Treutlein, Herbert R.
; TITLE OF INVENTION: CATALYTIC ANTIBODIES AND A METHOD OF
; TITLE OF INVENTION: PRODUCING SAME
; NUMBER OF SEQUENCES: 14
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: SCULLY, SCOTT, MURPHY & PRESSER
; STREET: 400 Garden City Plaza
; CITY: Garden City
; STATE: New York
; COUNTRY: United States of America
; ZIP: 11530

```



```

; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/160,567
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/828,741
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Digiglio, Frank S.
; REGISTRATION NUMBER: 31,346
; REFERENCE/DOCKET NUMBER: 10591
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (516) 742-4343
; TELEFAX: (516) 742-4366
; TELEX: 230 901 SANS UR
; INFORMATION FOR SEQ ID NO: 6:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 342 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-09-160-567-6

```

```

Query Match 27.7%; Score 618; DB 4; Length 342;
Best Local Similarity 75.9%; Pred. No. 8.4e-42;
Matches 126; Conservative 5; Mismatches 15; Indels 20; Gaps 1;

QY 16 SEEVTIKANLIFANGSTQTAEPKGTPEKATSEAYAYADTLKKDNGEYTVDVADKGYTLN 75
Db 174 SAEVTIKANLIFANGSTQTAEPKGTPEKATSEAYAYADTLKKDNGEYTVDVADKGYTLN 233

QY 76 IKFAGKEKTPPEE-----PKEEVTIKANLIYADGKTQTAEFKGTPE 115
Db 234 IKFAGKEKTPPEE-----PKEEVTIKANLIYADGKTQTAEFKGTPE 293

QY 116 EATAEAYRAYADALKKNGEYTVDVADKGYTLNKKFAGKEKTPPEPK 161
Db 294 KATSEAYAYADTLKKDNGEYTVDVADKGYTLNKKFAGKESAWRHPQ 339

```

```

RESULT 12
US-08-378-761A-19
; Sequence 19, Application US/08378761A
; Patent No. 5635384
; GENERAL INFORMATION:
; APPLICANT: WALSH, TERENCE A
; APPLICANT: HEY, TIMOTHY D
; APPLICANT: MORGAN, ALICE ER
; TITLE OF INVENTION: RIBOSOME-INACTIVATING PROTEINS, INACTIVE
; TITLE OF INVENTION: PRECURSOR FORMS THEREOF, A PROCESS FOR MAKING A METHOD OF
; TITLE OF INVENTION: USING
; NUMBER OF SEQUENCES: 81
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: ANDREA T. BORUCKI
; STREET: 9330 ZIONSVILLE ROAD
; CITY: INDIANAPOLIS
; STATE: IN
; COUNTRY: US
; ZIP: 46268
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/378,761A
; FILING DATE: 26-Jan-1995
; CLASSIFICATION: 435

```

```

; ATTORNEY/AGENT INFORMATION:
; NAME: BORUCKI, ANDREA T
; REGISTRATION NUMBER: 33651
; REFERENCE/DOCKET NUMBER: 38272B
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (317) 337-4846
; INFORMATION FOR SEQ ID NO: 19:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 489 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-378-761A-19

Query Match 27.7%; Score 618; DB 1; Length 489;
Best Local Similarity 40.8%; Pred. No. 1.4e-41;
Matches 175; Conservative 35; Mismatches 131; Indels 88; Gaps 16;

QY 37 EFKGTF-----EKATSEAYAYADTLKKDNGEYTVDV-ADKGYTLNKKFAGKEKTPPEPK 89
Db 38 DHKGIFQPVLPPEKVPPELWFYTE-LKTRTSSITLAIKMDNLVGVGFTPG-----87

QY 90 EEVTIKANLIYADGKTQTAEFKGTPEEATAEAYAYADALKKNGEYTVDVADKGYTLN 149
Db 88 -----GVWWEFGKDGDTLLLDGNPRWLGFGGRYQD-LIGNKGLTETVTMGRAMTRAVN 139

QY 150 FAGKEKTPPEKPEVTIKANLIYADGKTQTAEFKGTPEEATAEAYAYADLLAKENGKTYV 209
Db 140 DLAKKKKADPQ-----ADTKSKLVK-----LVWVCEGLRENTV-----SRTV 178

QY 210 DV---ADKGYTLNKKFAGKEKTPPEKPEVTIKANLIYADGKTQT-AEFGKTFEAEATAEA 265
Db 179 DAGFNSOHGVTLVLT-QGK---QVQKWDRIKAAFEWADHPAVIPDMQKLGKDKNEA 233

QY 266 YRVADLLAKENGKYTADLEDCGGYTIINIRPAGK---VDEKPEEP-MDTYKLIINGKTLKG 321
Db 234 ARTVALVKNT-----TAAATAGSKPEVIDASELTPAVTYYKLVINGKTLKG 281

QY 322 ETTTEAVDAATAEKVKQYANDNGVGEWYDDATKTFTVTE-----K 364
Db 282 ETTTEAVDAATAEKVKQYANDNGVGEWYDDATKTFTVTEKPEVIDASELTPAVTRSK 341

QY 365 PEVIDASELTPAVTYYKLVINGKTLKGTTTKAVDAETAETAEKAFKQYANDNGVGEWYDD 424
Db 342 PEVIDASELTPAVTYYKLVINGKTLKGTTTEAVDAATAEKVKQYANDNGVGEWYDD 401

QY 425 ATKFTVTE 433
Db 402 ATKFTVTE 410

RESULT 13
US-08-485-286-19
; Sequence 19, Application US/08485286
; Patent No. 5646026
; Patent No. 5646026 5646119
; GENERAL INFORMATION:
; APPLICANT: WALSH, TERENCE A
; APPLICANT: HEY, TIMOTHY D
; APPLICANT: MORGAN, ALICE ER
; TITLE OF INVENTION: RIBOSOME-INACTIVATING PROTEINS, INACTIVE
; TITLE OF INVENTION: PRECURSOR FORMS THEREOF, A PROCESS FOR MAKING A METHOD OF
; TITLE OF INVENTION: USING
; NUMBER OF SEQUENCES: 81
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: ANDREA T. BORUCKI
; STREET: 9330 ZIONSVILLE ROAD
; CITY: INDIANAPOLIS
; STATE: IN
; COUNTRY: US
; ZIP: 46268
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk

```

; COMPUTER: IBM PC compatible
 ; OPERATING SYSTEM: PC-DOS/MS-DOS
 ; SOFTWARE: PatentIn Release #1.0, Version #1.25
 ; CURRENT APPLICATION DATA:
 ; APPLICATION NUMBER: US/08/485,286
 ; FILING DATE:
 ; CLASSIFICATION: 435
 ; PRIOR APPLICATION DATA:
 ; APPLICATION NUMBER: US 08/378761
 ; FILING DATE: 26-JAN-1995
 ; ATTORNEY/AGENT INFORMATION:
 ; NAME: BORUCKI, ANDREA T
 ; REGISTRATION NUMBER: 38272B
 ; REFERENCE/DOCKET NUMBER: 38272B
 ; TELECOMMUNICATION INFORMATION:
 ; TELEPHONE: (317) 337-4846
 ; INFORMATION FOR SEQ ID NO: 19:
 ; SEQUENCE CHARACTERISTICS:
 ; LENGTH: 489 amino acids
 ; TYPE: amino acid
 ; TOPOLOGY: linear
 ; MOLECULE TYPE: protein
 ; US-08-485-286-19

Query Match 27.7%; Score 618; DB 1; Length 489;
 Best Local Similarity 40.8%; Pred. No. 1.4e-41;
 Matches 175; Conservative 35; Mismatches 131; Indels 88; Gaps 16;

QY 37 EFKGTF-----EKATSEAYAYADTLKNGEYTVDV-ADKGYTLNKKFAGKEKTPPEPK 89
 DB 38 DHKGFQFVLPPEKKVPELWFYFE-LKTRTSSITLAIRMDNLVGVFRTPG-----87
 QY 90 EEVTIKANLIYADGKTQTAEFGKTFEEATAEAYADALKKNGEYTVDVADKGYTLNKK 149
 DB 88 -----GVWWEFGKGDTHLLGDNPRWLGFGGRYQD-LIGNKGLGVTHMGRMTRAVN 139
 QY 150 FAGKEKTPPEPKKEVTIKANLIYADGKTQTAEFGKTFEEATAEAYADALKKNGEYTV 209
 DB 140 DLAKKKAAADPQ-----ADTKSKLVK---LVWVVCGLRENTV-----SRTV 178
 QY 210 DV---ADKGYTLNKKFAGKEKTPPEPKKEVTIKANLIYADGKTQT-AEFGKTFEATAEA 265
 DB 179 DAGFNSQHGVTLAVT-OGK---QVQKWDRIKSAAFADHPHTAVIPDMQKLGKDKNEA 233
 QY 266 YRVADLLAKNGKYTADLDGGYTVINIRFAGKK---VDEKPEEP-NDYTKLIILNGKTLKG 321
 DB 234 ARIVALVKNT-----TAAATAGSKPEVIDASELTPAVTTTKLVINGKTLKG 281
 QY 322 ETTEAVDAATAEKVKQYANDNGVDGEWYDDATKFTVTY-----K 364
 DB 282 ETTTEAVDAATAEKVKQYANDNGVDGEWYDDATKFTVTYKPEVIDASELTPAVTRSK 341
 QY 365 PEVIDASELTPAVTTTKLVINGKTLGETTTTKAVDAETAFAKQYANDNGVDGVTYDD 424
 DB 342 PEVIDASELTPAVTTTKLVINGKTLGETTTTEAVDAATAEKVKQYANDNGVDGEWYDD 401
 QY 425 ATKFTVTYE 433
 DB 402 ATKFTVTYE 410

RESULT 14
 US-08-828-741B-13
 ; Sequence 13, Application US/08828741B
 ; Patent No. 6043069
 ; GENERAL INFORMATION:
 ; APPLICANT: Koentgen, Frank
 ; APPLICANT: Suess, Gabriele M.
 ; APPLICANT: Tarlington, David M.
 ; APPLICANT: Treutlein, Herbert R.
 ; TITLE OF INVENTION: CATALYTIC ANTIBODIES AND A METHOD OF
 ; NUMBER OF SEQUENCES: 14

; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: SCULLY, SCOTT, MURPHY & PRESSER
 ; STREET: 400 Garden City Plaza
 ; CITY: Garden City
 ; STATE: New York
 ; COUNTRY: United States of America
 ; ZIP: 11530
 ; COMPUTER READABLE FORM:
 ; MEDIUM TYPE: Floppy disk
 ; COMPUTER: IBM PC compatible
 ; OPERATING SYSTEM: PC-DOS/MS-DOS
 ; SOFTWARE: PatentIn Release #1.0, Version #1.30
 ; CURRENT APPLICATION DATA:
 ; APPLICATION NUMBER: US/08/828,741B
 ; FILING DATE: 26-MAR-1997
 ; CLASSIFICATION: 435
 ; ATTORNEY/AGENT INFORMATION:
 ; NAME: Digiglio, Frank S.
 ; REGISTRATION NUMBER: 31,346
 ; REFERENCE/DOCKET NUMBER: 10591
 ; TELECOMMUNICATION INFORMATION:
 ; TELEPHONE: (516) 742-4343
 ; TELEFAX: (516) 742-4366
 ; TELEX: 230 901 SANS UR
 ; INFORMATION FOR SEQ ID NO: 13:
 ; SEQUENCE CHARACTERISTICS:
 ; LENGTH: 178 amino acids
 ; TYPE: amino acid
 ; TOPOLOGY: linear
 ; MOLECULE TYPE: protein
 ; US-08-828-741B-13

Query Match 27.5%; Score 615; DB 3; Length 178;
 Best Local Similarity 76.2%; Pred. No. 5.9e-42;
 Matches 125; Conservative 5; Mismatches 14; Indels 20; Gaps 1;

QY 18 EEVTIKANLIYADGKTQTAEFGKTFEEKATSEAYAYADTLKNGEYTVDVADKGYTLNKK 77
 DB 12 EEVTIKANLIYADGKTQTAEFGKTFEEKATSEAYAYADTLKNGEYTVDVADKGYTLNKK 71
 QY 78 FAGKEKTPPE-----PKEEVTIKANLIYADGKTQTAEFGKTFEEA 117
 DB 72 FAGKEATNTDGTDDYGLQINRWGGLTSABEVTIKANLIYADGKTQTAEFGKTFEEA 131
 QY 118 TAEAYAYADALKKNGEYTVDVADKGYTLNKKFAGKEKTPPEPK 161
 DB 132 TSEAYAYADTLKNGEYTVDVADKGYTLNKKFAGKESAWRHPQ 175

RESULT 15
 US-09-160-567-13
 ; Sequence 13, Application US/09160567
 ; Patent No. 6326179
 ; GENERAL INFORMATION:
 ; APPLICANT: Koentgen, Frank
 ; APPLICANT: Suess, Gabriele M.
 ; APPLICANT: Tarlington, David M.
 ; APPLICANT: Treutlein, Herbert R.
 ; TITLE OF INVENTION: CATALYTIC ANTIBODIES AND A METHOD OF
 ; NUMBER OF SEQUENCES: 14
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: SCULLY, SCOTT, MURPHY & PRESSER
 ; STREET: 400 Garden City Plaza
 ; CITY: Garden City
 ; STATE: New York
 ; COUNTRY: United States of America
 ; ZIP: 11530
 ; COMPUTER READABLE FORM:
 ; MEDIUM TYPE: Floppy disk
 ; COMPUTER: IBM PC compatible
 ; OPERATING SYSTEM: PC-DOS/MS-DOS
 ; SOFTWARE: PatentIn Release #1.0, Version #1.30

```

; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/160,567
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA: 08/828,741
; APPLICATION NUMBER: 08/828,741
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Digiglio, Frank S.
; REGISTRATION NUMBER: 31,346
; REFERENCE/DOCKET NUMBER: 10591
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (516) 742-4343
; TELEFAX: (516) 742-4366
; TELEX: 230 901 SANS UR
; INFORMATION FOR SEQ ID NO: 13:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 178 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-09-160-567-13

Query Match      27.5%; Score 615; DB 4; Length 178;
Best Local Similarity 76.2%; Pred. No. 5.9e-42;
Matches 125; Conservative 5; Mismatches 14; Indels 20; Gaps 1;

QY 18 EEVTIKANLIFANGSTQTAEFKGTFFKATSEAYAYADTLKKDNGEYTVDVADKGYTLNIK 77
Db 12 EEVTIKANLIFANGSTQTAEFKGTFFKATSEAYAYADTLKKDNGEYTVDVADKGYTLNIK 71
QY 78 FAGKEKTPPEE-----PKEEVTIKANLIYADGKTQTAEFKGTPEEA 117
Db 72 FAGKEATNRNTDGSYGIILQINSRWGGLTSAEEVTIKANLIFANGSTQTAEFKGTPEKA 131
QY 118 TAEAYRYADALKKDNGEYTVDVADKGYTLNIKFKAGKEKTPPEPK 161
Db 132 TSEAYAYADTLKKDNGEYTVDVADKGYTLNIKFKAGKESAWRHPQ 175

```

Search completed: October 29, 2002, 09:31:57
Job time : 16.5074 secs

THIS PAGE BLANK (uspto)

GenCore version 5.1.3
Copyright (c) 1993 - 2002 Compugen Ltd.

OM protein - protein search, using sw model

Run on: October 29, 2002, 09:25:06 : Search time 34.0622 Seconds
(without alignments)
1415.233 Million cell updates/sec

Title: US-08-325-278B-3

Perfect score: 2235

Sequence: 1 AVENKEETPETDSEEV.....GVDGVWYDADTKFTVTTEM 434

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 747574 seqs, 11107396 residues

Total number of hits satisfying chosen parameters: 747574

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

A_Geneseq_032802:*
1: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1980.DAT:*
2: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1981.DAT:*
3: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1982.DAT:*
4: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1983.DAT:*
5: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1984.DAT:*
6: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1985.DAT:*
7: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1986.DAT:*
8: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1987.DAT:*
9: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1988.DAT:*
10: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1989.DAT:*
11: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1990.DAT:*
12: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1991.DAT:*
13: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1992.DAT:*
14: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1993.DAT:*
15: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1994.DAT:*
16: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1995.DAT:*
17: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1996.DAT:*
18: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1997.DAT:*
19: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1998.DAT:*
20: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1999.DAT:*
21: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA2000.DAT:*
22: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA2001.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	ID	Description
1	2235	100.0	434	14	Sequence encoding
2	1565	70.0	305	14	Immunoglobulin lig
3	1550	69.4	467	22	Amino acid sequenc
4	1372	61.4	367	21	Expression vector
5	1263.5	56.5	1027	14	Protein L. Peptoc
6	1263.5	56.5	1027	14	Protein L. Peptoc
7	1216	54.4	291	14	Immunoglobulin bin
8	706.5	31.6	664	16	S. dysgalactiae MI
9	702	31.4	593	15	Streptococcus Prot
10	701.5	31.4	594	12	Streptococcus GX78
11	692.5	31.0	480	8	Sequence of polype

12	692	31.0	593	11	AA07014	Protein G variant
13	683	30.6	448	8	AA070493	Protein G. Strept
14	683	30.6	448	10	AA095030	Protein G. Strept
15	683	30.6	448	11	AA07013	Protein G variant
16	683	30.6	448	12	AA01004	Streptococcus GX78
17	683	30.6	448	15	AA053290	Streptococcus Prot
18	675	30.2	269	15	AA053295	IgG-binding Strept
19	674	30.2	265	12	AA010011	Type 4 GX7809 prot
20	674	30.2	265	15	AA053294	IgG-binding Strept
21	670	30.0	185	21	AA057610	Streptococcus stra
22	665	29.8	235	10	AA094785	Protein G variant.
23	665	29.8	235	11	AA07004	Protein G variant.
24	663	29.7	144	10	AA094784	Protein G variant.
25	663	29.7	144	11	AA07003	Protein G variant.
26	663	29.7	144	12	AA010009	Type 3 GX7809 prot
27	663	29.7	144	15	AA053293	IgG-binding Strept
28	663	29.7	164	10	AA094787	Protein G variant.
29	663	29.7	164	11	AA07011	Protein G variant.
30	663	29.7	164	12	AA010015	Type 9 GX7809 prot
31	663	29.7	164	15	AA053299	IgG-binding Strept
32	663	29.7	199	10	AA094783	Protein G variant.
33	663	29.7	199	11	AA07002	Protein G variant.
34	663	29.7	199	12	AA010007	Type 2 GX7809 prot
35	663	29.7	199	15	AA053292	IgG-binding Strept
36	663	29.7	201	12	AA010013	Type 8 GX7809 prot
37	663	29.7	201	15	AA053298	IgG-binding Strept
38	658	29.4	148	10	AA094782	Protein G variant.
39	658	29.4	148	11	AA07001	Protein G variant.
40	658	29.4	148	12	AA010008	Type 1 GX7809 prot
41	658	29.4	148	15	AA053291	IgG-binding Strept
42	657	29.4	201	11	AA07010	Protein G variant.
43	656	29.4	269	10	AA094788	Protein G variant.
44	648	29.0	269	11	AA07006	Protein G variant.
45	646	28.9	201	10	AA094790	Protein G variant.

ALIGNMENTS

RESULT 1
AA042994

ID AA042994 standard; Protein: 434 AA.

XX AC AA042994:

XX DT 16-MAY-1994 (first entry)

XX Sequence encoding immunoglobulin light chain binding protein.

DE Immunoglobulin; light chain; binding; identification; purification;
KW separation; ss.

XX OS E. coli L392/pHDLG, DSM 7055.

XX Key Location/Qualifiers

FH Domain 5..305

FT /label= B1 immunoglobulin light chain binding domain.

FT Domain 81..305

FT /label= B2 immunoglobulin light chain binding domain.

FT Domain 153..305

FT /label= B3 immunoglobulin light chain binding domain.

FT Domain 225..305

FT /label= B4 immunoglobulin light chain binding domain.

FT Domain 297..305

FT /label= B5 immunoglobulin light chain binding domain.

FT Domain 309..434

FT /label= C1 immunoglobulin heavy chain binding domain.

```

FT Domain 364..434
FT /label= D intermediate immunoglobulin heavy
FT chain binding domain.
FT
FT Domain 379..434
FT /label= C2 immunoglobulin heavy chain binding
FT domain.
XX
XX WO9322342-A.
XX
XX 11-NOV-1993.
XX
XX 28-APR-1993; 93WO-SE00375.
XX
XX 28-APR-1992; 92SE-0001331.
XX
XX (HIGH-) HIGHTECH RECEPTOR AB.
XX
XX Bjoerck L, Sjoerbring U;
XX
XX WPI; 1993-368722/46.
XX P-PSDB; AAR42994.
XX
XX New protein L binding light chains of all immunoglobulin classes
XX - for binding purifying and identifying immunoglobulin, also
XX related DNA, vectors and host cells
XX
XX Claim 6; Page 39-40; 71pp; English.
XX
XX Protein L (AAR42993) is capable of binding to immunoglobulin G light
XX chains. It is useful for binding, separating (purifying) and
XX identifying immunoglobulin and for removing immunoglobulin molecules
XX from serum. This is the coding sequence of one hybrid protein of
XX the L protein. The hybrid proteins can bind all human
XX immunoglobulin classes and many immunoglobulins from other species.
XX They are highly soluble and retain their binding activity at high
XX temperatures over a pH range of 3-10. They can be immobilised
XX without loss of activity.
XX
XX Sequence 434 AA;
XX
Query Match 100.0%; Score 2235; DB 14; Length 434;
Best Local Similarity 100.0%; Pred. No. 1.1e-123;
Matches 434; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 AVENKEETPETDSEEVITKANLIFANGSTQTAEFKGTSEAYAYADTLKKN 60
DB 1 AVENKEETPETDSEEVITKANLIFANGSTQTAEFKGTSEAYAYADTLKKN 60
QY 61 GEYTVADVADKGYTLNKFAGKEKTPPEPKKEVITKANLIYADGKTQTAEPKGFEEATAE 120
DB 61 GEYTVADVADKGYTLNKFAGKEKTPPEPKKEVITKANLIYADGKTQTAEPKGFEEATAE 120
QY 121 AYRYADALKDNGEYTVADVADKGYTLNKFAGKEKTPPEPKKEVITKANLIYADGKTQTA 180
DB 121 AYRYADALKDNGEYTVADVADKGYTLNKFAGKEKTPPEPKKEVITKANLIYADGKTQTA 180
QY 181 EFKGTPEATAEAYRYADLLAKENGKGYTVADVADKGYTLNKFAGKEKTPPEPKKEVITKA 240
DB 181 EFKGTPEATAEAYRYADLLAKENGKGYTVADVADKGYTLNKFAGKEKTPPEPKKEVITKA 240
QY 241 NLIYADGKTQTAEPKGFEEATAEAYRYADLLAKENGKGYTVADLEGGYTTINIRFAGKKVD 300
DB 241 NLIYADGKTQTAEPKGFEEATAEAYRYADLLAKENGKGYTVADLEGGYTTINIRFAGKKVD 300
QY 301 EKPEPMDTVKLLINGTKLGETTTEAVDAATAEAKVFKQYANDNGVDGEWYDDATKFTT 360
DB 301 EKPEPMDTVKLLINGTKLGETTTEAVDAATAEAKVFKQYANDNGVDGEWYDDATKFTT 360
QY 361 VTEKPEVIDASELTPATVYTKLVINGTKLGETTTRAVDAATAEAKVFKQYANDNGVDGVW 420
DB 361 VTEKPEVIDASELTPATVYTKLVINGTKLGETTTRAVDAATAEAKVFKQYANDNGVDGVW 420
QY 421 TYDDATKFTVTTEM 434

```

```

Db 421 TYDDATKFTVTTEM 434
|||||
RESULT 2
AAR42993
ID AAR42993 standard; Protein; 305 AA.
XX
XX AAR42993;
XX
XX 16-MAY-1994 (first entry)
XX
XX Immunoglobulin light chain binding protein (protein L).
XX Immunoglobulin; light chain; binding; identification; purification;
XX separation.
XX
XX E. coli LE392/pHDL, DSM 7054.
XX
XX Key Location/Qualifiers
XX Domain 5..305
XX /label= B1 immunoglobulin light chain binding
XX domain.
XX Domain 81..305
XX /label= B2 immunoglobulin light chain binding
XX domain.
XX Domain 153..305
XX /label= B3 immunoglobulin light chain binding
XX domain.
XX Domain 225..305
XX /label= B4 immunoglobulin light chain binding
XX domain.
XX Domain 297..305
XX /label= B5 immunoglobulin light chain binding
XX domain.
XX WO9322342-A.
XX
XX 11-NOV-1993.
XX
XX 28-APR-1993; 93WO-SE00375.
XX
XX 28-APR-1992; 92SE-0001331.
XX
XX (HIGH-) HIGHTECH RECEPTOR AB.
XX
XX Bjoerck L, Sjoerbring U;
XX
XX WPI; 1993-368722/46.
XX N-PSDB; AAQ50452.
XX
XX New protein L binding light chains of all immunoglobulin classes
XX - for binding purifying and identifying immunoglobulin, also
XX related DNA, vectors and host cells
XX
XX Claim 1; Page 36; 71pp; English.
XX
XX The protein (protein L) is capable of binding to immunoglobulin G
XX light chains. It is useful for binding, separating (purifying) and
XX identifying immunoglobulin and for removing immunoglobulin molecules
XX from serum. Hybrid proteins of the L protein can bind all human
XX immunoglobulin classes and many immunoglobulins from other species.
XX They are highly soluble and retain their binding activity at high
XX temperatures over a pH range of 3-10. They can be immobilised
XX without loss of activity.
XX
XX Sequence 305 AA;
XX
Query Match 70.0%; Score 1565; DB 14; Length 305;
Best Local Similarity 100.0%; Pred. No. 1.4e-84;
Matches 305; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 AVENKEETPETDSEEVITKANLIFANGSTQTAEFKGTSEAYAYADTLKKN 60

```

|||||
Db 1 AVENKEETPTDSEEEVTKANLIFANGSTQTAEFKGTFEKATSEAYADTLKKN 60
QY 61 GEYTVADVADKGYTLNKFAGKEKTPPEPKKEVTKANLIYADGKTQTAEFKGTFEATAE 120
Db 61 GEYTVADVADKGYTLNKFAGKEKTPPEPKKEVTKANLIYADGKTQTAEFKGTFEATAE 120
QY 121 AYRYADALKKDNGEYTVADVADKGYTLNKFAGKEKTPPEPKKEVTKANLIYADGKTQTA 180
Db 121 AYRYADALKKDNGEYTVADVADKGYTLNKFAGKEKTPPEPKKEVTKANLIYADGKTQTA 180
QY 181 EFKGTFEATAEAYRYADLLAKENGYTVADVADKGYTLNKFAGKEKTPPEPKKEVTKA 240
Db 181 EFKGTFEATAEAYRYADLLAKENGYTVADVADKGYTLNKFAGKEKTPPEPKKEVTKA 240
QY 241 NLIYADGKTQTAEFKGTFAEATAEAYRYADLLAKENGYTADLEDDGGYTNIRFAGKKVD 300
Db 241 NLIYADGKTQTAEFKGTFAEATAEAYRYADLLAKENGYTADLEDDGGYTNIRFAGKKVD 300
QY 301 EKPEE 305
Db 301 EKPEE 305
RESULT 3
AAB31372
ID AAB31372 standard; Protein; 467 AA.
AC AAB31372;
XX
XX
DT 20-APR-2001 (first entry)
XX
XX Amino acid sequence of protein L/CBD cex/ER retaining peptide fusion.
XX
XX Protein production; food processing; protein antibiotic; feed enzyme;
KW protein L: CBD cex protein; cell signal peptide.
XX
XX Synthetic.
OS
XX WO20007174-A1.
PN
XX
XX 21-DEC-2000.
PF
XX 07-JUN-2000; 2000WO-IL00330.
XX
XX 10-JUN-1999; 99US-0329234.
PR
XX (CBDT-) CBD TECHNOLOGIES LTD.
PA (YISS) YISSUM RES DEV CO HEBREW UNIV JERUSALEM.
XX
XX Shani Z, Shoseyov O;
PI
XX WPI; 2001-112219/12.
DR N-PSDB; AAF24730.
XX
XX Expressing and isolating recombinant protein in a plant, useful for
PT producing large quantities of recombinant proteins, by expressing a
PT fusion protein including a cellulose binding peptide fused to a
PT recombinant protein
XX
XX Example; Fig 2a; 87pp; English.
PS
XX
XX The specification describes a method for expressing and isolating
CC a recombinant protein in a plant. The method comprising expressing a
CC fusion protein including the recombinant protein and a cellulose
CC binding peptide fused to it, where the fusion protein is
CC compartmentalised and sequestered within plant cells, plant
CC tissue or cultured plant cells. The method is useful for obtaining large
CC quantities of the recombinant proteins and protein products in a simple
CC and cost-effective manner. Recombinant proteins may be used commercially,
CC such as in the food processing industry, e.g. glucosylases and glucose
CC isomerases are used for converting starch to high fructose corn syrup,
CC proteinases for the hydrolysis of high molecular weight proteins and in

CC manufacturing leather or alcoholic beverages, pectinesterases for
CC pectin hydrolysis in food industry, lipases for cleaving ester linkage
CC in triglycerides, and for effluent treatment. The recombinant proteins
CC may further be used to produce protein antibiotics, which can be used
CC in healing processes, and to produce animal feed enzymes. The present
CC sequence represents a fusion protein of the invention, and comprises a
CC fusion of a cell signal peptide, protein L, CBD cex, and an endoplasmic
CC reticulum retaining peptide.
XX
SQ Sequence 467 AA;
Query Match 69.4%; Score 1550; DB 22; Length 467;
Best Local Similarity 79.2%; Pred. No. 1.8e-83;
Matches 328; Conservative 3; Mismatches 43; Indels 40; Gaps 7;
QY 1 AVENKEETPTDSEEEVTKANLIFANGSTQTAEFKGTFEKATSEAYADTLKKN 60
Db 40 AVENKEETPTDSEEEVTKANLIFANGSTQTAEFKGTFEKATSEAYADTLKKN 99
QY 61 GEYTVADVADKGYTLNKFAGKEKTPPEPKKEVTKANLIYADGKTQTAEFKGTFEATAE 120
Db 100 GEYTVADVADKGYTLNKFAGKEKTPPEPKKEVTKANLIYADGKTQTAEFKGTFEATAE 159
QY 121 AYRYADALKKDNGEYTVADVADKGYTLNKFAGKEKTPPEPKKEVTKANLIYADGKTQTA 180
Db 160 AYRYADALKKDNGEYTVADVADKGYTLNKFAGKEKTPPEPKKEVTKANLIYADGKTQTA 219
QY 181 EFKGTFEATAEAYRYADLLAKENGYTVADVADKGYTLNKFAGKEKTPPEPKKEVTK 239
Db 220 EFKGTFEATAEAYRYADLLAKENGYTVADVADKGYTLNKFAGKEKTPPEPKKEVTK 279
QY 240 ANLIYADGKTQTAEFKGTFAEATAEAYRYADLLAKENGYTADLEDDGGYTNIRFAGKKV 299
Db 280 ANLIYADGKTQTAEFKGTFAEATAEAYRYADLLAKENGYTADLEDDGGYTNIRFAGKKV 339
QY 300 DEKPEEPMQYTKLILNGTKLKGTTTTEAVDAA-----TAEKVFQYANDN 344
Db 340 DEKPEGIPPT-----PTPTSASGPAGCQVLWGVNQMNTGTAOVTVKN-TCGA 386
QY 345 GVDGEWYDADTKTFTVTEKPEVIDASELT-----PAVTYKLVINGTKLKG 393
Db 387 PVDG-WTL-----TFSPSGQQVQAWSTVOTSGSAVTVRNAPWNGNIPAGGT 434
RESULT 4
AAB10432
ID AAB10432 standard; Protein; 367 AA.
XX
XX AAB10432;
DT 01-DEC-2000 (first entry)
XX
XX Expression vector pSEX11L4 protein G.
DE
XX
XX Expression vector; antibody binding protein; monoclonal antibody; Neo-R;
KW B lymphocyte; myeloma cell; hybridoma; protein G; bla protein.
XX
XX Synthetic.
OS
XX DE19900635-A1.
PN
XX 13-JUL-2000.
PD
XX
XX 11-JAN-1999; 99DE-1000635.
XX
XX 11-JAN-1999; 99DE-1000635.
XX
XX (DEKR-) DEUT KREBSFORSCHUNGSZENTRUM.
XX
XX Breitling F, Poustka A, Moldenhauer G;
PI WPI; 2000-499832/45.
XX
XX N-PSDB; AAA71428.
DR

XX Selecting monoclonal antibodies, by expressing them on the surface of
PT hybridomas attached to antibody-binding protein, then reaction with
PT antibody library -
XX
XX Claim 16; Fig 1; 22pp; German.
XX
XX This invention describes a novel method for the selection of monoclonal
CC antibodies (MAB) which comprises (i) fusing B lymphocytes with myeloma
CC cells to produce antibody-producing hybridomas such that the antibodies
CC are presented at the surface of the hybridomas by an antibody-binding
CC protein (I); and (ii) binding the antibody to antigens (Ag). The
CC invention also describes antibody-binding proteins (I) that comprise a
CC combination of the signal peptide of a murine immunoglobulin (Ig) kappa
CC chain or a murine MHC (major histocompatibility complex) Class I k(k)
CC molecule; an antibody-binding site of proteins A, G, L or LG, and the
CC transmembrane domain of PDGFR (platelet-derived growth factor receptor)
CC or CD52. The method is used to select MAB with specificity for particular
CC antigens. MAB can be selected without separate culture of hybridomas,
CC and selection can be made against many antigens in a library, optionally
CC on the basis of strength of affinity for a particular antigen. Complex
CC mixtures of hybridomas can be used for selection, reducing the time and
CC cost involved in MAB selection. This sequence represents the protein G,
CC contained in the expression vector pSEXil14 which contains the protein G,
CC Neo-R and the bla protein described in the method of the invention.
XX
XX Sequence 367 AA;

Query Match 61.4%; Score 1372; DB 21; Length 367;
Best Local Similarity 82.4%; Pred. No. 3.6e-73;
Matches 271; Conservative 19; Mismatches 29; Indels 10; Gaps 3;
QY 5 KEETPEPTDSSEEVITKANLIFANGSTOTAEFGTFEKATSEAYAYADTLKKDNGEYT 64
DB 25 KEKTPPEP-----KEEVITKANLIYADGKTQTAEFGTFEATAEAYADALAKKDNGEYT 80
QY 65 VDADKGYTLNIFAGKEKTPPEKPEVITKANLIYADGKTQTAEFGTFEATAEAYRY 124
DB 81 VDADKGYTLNIFAGKEKTPPEKPEVITKANLIYADGKTQTAEFGTFEATAEAYRY 140
QY 125 ADALAKKDNGEYTVDVADKGYTLNIFAGKEKTPPEKPEVITKANLIYADGKTQTAEFGK 184
DB 141 ADALAKKDNGEYTVDVADKGYTLNIFAGKEKTPPEKPEVITKANLIYADGKTQTAEFGK 200
QY 185 TFEATAEAYRYADLAKENGKTYTVDVADKGYTLNIFAGKEKTPPEKPEVITKANLIY 244
DB 201 TFEATAEAYRYADLAKENGKTYTVDVADKGYTLNIFAGKEKTPPEKPEVITKANLIY 260
QY 245 ADGKTQTAEFGKTFEATAEAYRYADLAKENGKTYTVDVADKGYTLNIFAGKTVDEKPE 304
DB 261 ADGKTQTAEFGKTFEATAEAYRYADLAKKDNGEYTVDVADKGYTLNIFAGAAEQKLI 320
QY 305 EPMDTYKLIINGTKLGETTTEAVDAATA 333
DB 321 SEED-----LNG-AVDQNDTSQTSPSA 343

RESULT 5
AAR42203
ID AAR42203 standard; Protein; 1027 AA.
XX
AC AAR42203;
XX
DT 18-MAY-1994 (first entry)
XX
DE Protein L.
XX
KW Peptide; immunoglobulin; binding; analysis; purification; ELISA;
KW enzyme linked immunoabsorbant assay.
XX
OS Peptococcus magnus.
XX
FH Key Location/Qualifiers

Peptide 36..59
/label= Signal sequence.
Protein 60..968
/label= Mature protein L.
XX
XX WO9322439-A.
XX
XX 11-NOV-1993.
XX
XX 07-MAY-1993; 93WO-CB00950.
XX
XX 07-MAY-1992; 92GB-0009804.
XX
XX 24-DEC-1992; 92GB-0026928.
XX
XX (PUBL-) PUBLIC HEALTH LAB SERVICE BOARD.
XX
XX Atkinson A, Duggleby CJ, Murphy JP, Trowern AR;
XX
XX WPI; 1993-368798/46.
XX
XX N-PSDB; AAQ50946.
XX
XX New immunoglobulin binding proteins derived from Protein L -
XX which bind immunoglobulin kappa light chains but not albumin or
XX cell walls
XX
XX Disclosure; Figure 1; 28pp; English.
XX
XX The synthetic immunoglobulin binding proteins derived from protein
XX L correspond to the repeated sequences in protein L which bind
XX immunoglobulin kappa light chains. They can be used in protein
XX analysis, purification procedures and other biochemical processes e.
XX g. ELISA. The synthetic molecules are of particular advantage if
XX they are free of regions in protein L which exhibit albumin and cell
XX wall binding.
XX
XX Sequence 1027 AA;
QY Query Match 56.5%; Score 1263.5; DB 14; Length 1027;
Best Local Similarity 58.8%; Pred. No. 2.7e-66;
Matches 281; Conservative 36; Mismatches 90; Indels 71; Gaps 12;
QY 2 VENKEETPEPTDSSEEVITKANLIFANGSTOTAEFGTFEKATSEAYAYADTLKKDNG 61
DB 276 MERKLESEKTEPE-----PEEVITKANLIFADGSTQNAEFGTFEAKAVSDAYADALAKKDNG 333
QY 62 EYTVDVADKGYTLNIFAGKEKTPPEKPEVITKANLIYADGKTQTAEFGTFEATAE 121
DB 334 EYTVDVADKGLTNIFAGKKEKPEKPEVITKANLIYADGKTQTAEFGTFEATAE 393
QY 122 YRYADALAKKDNGEYTVDVADKGYTLNIFAGKEKTPPEKPEVITKANLIYADGKTQT 179
DB 394 YAYADLAKENGKTYTVDVADKGYTLNIFAGKEKTPPEKPEVITKANLIYADGKTQT 453
QY 180 AEFKGTFEATAEAYRYADLAKENGKTYTVDVADKGYTLNIFAGKEKTPPEKPEVIT 237
DB 454 AEFKGTFEATAEAYRYADLAKENGKTYTVDVADKGYTLNIFAGKEKTPPEKPEVIT 513
QY 238 IKANLIYADGKTQTAEFGTFEATAEAYRYADLAKENGKTYTVDVADKGYTLNIFAGK 297
DB 514 IKVNLIFADGKTQTAEFGTFEATAEAYRYADLAKENGKTYTVDVADKGYTLNIFAGK 573
QY 298 K-----VDE-----KPEEPM-----DTY-KLILNGKTLKG----- 321
DB 574 EOPGNPGITIDENLLKNAKEAEIKELKEAGITSDLYFSLINKAKTVEGVEALKNEILKA 633
QY 322 -----ET-----TTEAVDAATAEAKVFKQYAND-----NGVDGEYTDATFTVTTEK 364
DB 634 HAGEETPELKDGYATYEEAEAAKAEALKNDDVNNAYEIVQAGDGRYYY--VLKIEVADEE 691
QY 365 PEVIDASELTPAVTYTKLVINGKTLKGETTTKAVDAETAEAKFKQYANDNGVDGVVTY 422
DB 692 EPGEDTPEVOEGYATYE-----EAAAKAEALKEKDVNNAYEVQAGDGRYYY 739

QY 130 KNGEYTVDVADKGYTLNIFKAGKE--KTPEPKKEEVTIKANLIYADGKTQTAEFKGTFE 187
 Db 120 KNGEYTDADGNGTINIFKAGKETPEPEPKKEEVTIKVNLIFADGKIQTAEFKGTFE 179
 QY 188 EATAEAYRYADLLAKENGKTYTVADVADKGYTLNIFKAGKE--KTPEPKKEEVTIKANLIYA 245
 Db 180 EATAKAYAYANLLAKENGKTYTVADVADKGYTLNIFKAGKETPEPEPKKEEVTIKVNLIFA 239
 QY 246 DGKTQTAEFKGTAEATAEAYRYADLLAKENGKTYTVADVADKGYTLNIFKAGKE--KTPEPKKEEVTIKANLIYA 297
 Db 240 DGKTQTAEFKGTAEATAEAYRYADLLAKENGKTYTVADVADKGYTLNIFKAGKE--KTPEPKKEEVTIKANLIYA 291

RESULT 8
 AAR71929
 ID AAR71929 standard; Protein; 664 AA.
 XX AAR71929;
 XX 22-SEP-1995 (first entry)
 DE S. dysgalactiae MIG.
 XX MIG; fast alpha-2-macroglobulin binding protein; FAM;
 KW plasma proteinase-inhibitor binding protein.
 XX Streptococcus dysgalactiae strain SC1.
 XX Key Location/Qualifiers
 FT Peptide 1..30
 FT /label= Sig_peptide
 FT Domain 209..278
 FT /label= IgG1
 FT /note= "IgG binding domain 1"
 FT Domain 279..348
 FT /label= IgG2
 FT /note= "IgG binding domain 2"
 FT Domain 349..418
 FT /label= IgG3
 FT /note= "IgG binding domain 3"
 FT Domain 419..488
 FT /label= IgG4
 FT /note= "IgG binding domain 4"
 FT Domain 489..558
 FT /label= IgG5
 FT /note= "IgG binding domain 5"
 FT Region 559..638
 FT /note= "cell wall spanning region"
 FT Peptide 630..635
 FT /note= "putative wall anchoring motif"
 FT Region 639..664
 FT /note= "membrane spanning domain"

XX W09507296-A.
 XX 16-MAR-1995.
 XX 06-SEP-1994; 94WO-SE00826.
 XX 06-SEP-1993; 93SE-0002855.
 XX (GUSS/) GUSS B.
 XX (JONS/) JONSSON H.
 XX (LIND/) LINDBERG M.
 XX (MUEL/) MUELLER H.
 XX (RANT/) RANTAMAKI L K.
 XX Guss B, Jonsson H, Lindberg M, Mueller H, Rantamaki LK;
 XX WPI; 1995-123382/16.
 XX N-PSDB; AAQ89196.
 XX DNA encoding fast alpha 2-macro:globulin-binding proteins - used

PT to obtain prods. for sepn., detection or quantification or for
 binding inhibition
 XX Disclosure; Page 31; 50pp; English.
 XX A phage lambda GEM-11 library of S. dysgalactiae SC1 DNA was analyzed
 CC for fast-acting alpha-2-macroglobulin (FAM)- and IgG-binding
 CC activity. One clone, which expressed both activities, was
 CC analyzed to obtain DNA encoding the FAM-binding protein, MIG
 CC (AAR71929). The mig gene is given in AAQ89196.
 XX Sequence 664 AA:
 SQ
 Query Match 31.6%; Score 706.5; DB 16; Length 664;
 Best Local Similarity 43.9%; Pred. No. 7.4e-34;
 Matches 192; Conservative 51; Mismatches 143; Indels 51; Gaps 17;
 QY 24 ANLIFANGSTQTAETFEKATSEAYAYADTLK-KDNGEYTVDVADKGYTL--NIKFA- 79
 Db 148 ANEIVNNSDAYTAESIPLYKLINDAY---DVLESKDYSKY--DSQDKVNNLADQLRDV 202
 QY 80 -----GKEKTPPEPKKEEVTIKANLIYADGKTQTAEFKGTFEATAEA--YRYA 125
 Db 203 QAVQLEAPTVIDAPELTPALTYTKLVKGNTF--SGETTK----AIDTATAKEEFKQYA 256
 QY 126 DALKDKNGEYTVDVADKGYTLNIFKFA---GKEKTPPEPKKEEVTIKANLIYADGKTQTAEF 182
 Db 257 TANNVD-GEWSYDDATKTFVTVEKPAVIDAPELTPALTYTKLVKGNTF--SGETTT--- 310
 QY 183 KGTFEATAEAYRYADLLAKENGKTYTVADVADKGYTLNIFKFA---GKEKTPPEPKKEEVTIK 239
 Db 311 KAVDAETAFAKQYATANNVDGEWSYDDATKTFVTVEKPAVIDAPELTPALTYTKLVK 370
 QY 240 ANLIYADGKTQTAEFKGTAEATAEAYRYADLLAKENGKTYTVADVADKGYTLNIFKAGKV 299
 Db 371 GNTF--SGETTTKADAATAEAEKQYATANGV---DGESYDDATKTFVTVEKPA---V 422
 QY 300 DEKPE--EPMDTYKLILNGKLTGETTTEAVDAATAEKFVKQYANGVDGWTYDDATK 357
 Db 423 IDAPELTPALTYTKLVKGNTFSGETTTKAVDAETAFAKQYANGVGEWSYDDATK 482
 QY 358 TFTVTEKPEVIDASELTPAVTVTKLVINGKTLKGETTTTAVDAETAFAKQYANGVD 417
 Db 483 TFTVTEKPAVIDAPELTPALTYTKLVINGKTLKGETTTTAVDAETAFAKQYANGVD 542
 QY 418 GWTYDDATKTFVTTEM 434
 Db 543 GWTYDDATKTFVTTEM 559
 RESULT 9
 AAR62944
 ID AAR62944 standard; Protein; 593 AA.
 XX AAR62944;
 XX 10-JAN-1995 (first entry)
 XX Streptococcus Protein G derived from strain GX7805.
 DE Streptococcus Protein G; variant; IgG binding activity;
 KW immunoglobulin; Lancefield Group G; bacterial Fc receptor.
 XX Streptococcus sp. GX7805.
 XX Key Location/Qualifiers
 FH Domain 106..140
 FT /label= A1
 FT Region 141..178
 FT /label= a1
 FT /note= "linking region"
 FT Domain 179..215
 FT /label= A2

FT Region 216..253
FT /label= a2
FT /note= "linking region"
FT Domain 254..290
FT /label= A3
FT Misc-difference 269
FT /note= "corresponds to GGA codon"
FT Active-site 303..357
FT /label= B1
FT /note= "Involved in IgG binding activity"
FT Region 358..372
FT /label= b
FT /note= "linking region"
FT Active-site 373..427
FT /label= B3
FT /note= "Involved in IgG binding activity"
FT Misc-difference 376
FT /note= "corresponds to CCT codon"
FT Region 428..442
FT /label= b
FT /note= "linking region"
FT Active-site 443..497
FT /label= B2
FT /note= "Involved in IgG binding activity"
FT Misc-difference 466
FT /note= "corresponds to ACT codon"
FT Region 531..535
FT /label= C1
FT Region 536..540
FT /label= C2
FT Region 541..545
FT /label= C3
FT Region 546..550
FT /label= C4
FT Region 551..555
FT /label= C5
FT Misc-difference 592
FT /note= "corresponds to GAA codon"
XX
XX US5312901-A.
XX
XX 17-MAY-1994.
XX
XX 14-FEB-1986; 86US-0829354.
XX
XX 14-FEB-1986; 86US-0829354.
XX 23-APR-1986; 86US-0854887.
XX 19-JUN-1987; 87US-0063959.
XX 20-JUN-1988; 88US-0209236.
XX 19-JUN-1990; 90US-0540169.
XX 21-APR-1992; 92US-0871539.
XX
XX (PHAA) PHARMACIA LKB BIOTECHNOLOGY AB.
XX
XX Fahnstock SR;
XX
XX WPI; 1994-159179/19.
XX N-PSDB; AAQ75036.
XX
XX
XX New recombinant streptococcal protein G variants - useful for
XX antibody detection and purification and for therapy
XX
XX Example 5; Fig 9; 48pp; English.
XX
XX A 2.4kb HindIII fragment containing the entire coding sequence for
XX Protein G was isolated from Streptococcus GX7805 using the 1.9kb
XX Protein G coding sequence from Streptococcus GX7809. The protein G
XX has IgG-binding activity which has been localised to the B repeating
XX structure. Streptococcal Protein G variants comprising the B domains
XX are claimed.
XX
XX Sequence 593 AA;

Query Match 31.4%; Score 702; DB 15; Length 593;
Best Local Similarity 48.1%; Pred. No. 1.2e-33;
Matches 204; Conservative 28; Mismatches 132; Indels 60; Gaps 17;
QY 45 ATSEAYAYADTLKKDNGEYTVDVADKGYTLNIRFAGKEKTPPEPKBEVTIKANLIYA-DG 103
DB 101 ADALAKAKADALKEFN-KYGVSDYYKNLINNAKTVEGVKDLQAVVESAKKARISEATDG 159
QY 104 -----KTQTAEFKGTPEEATAEAYRYADALKDNGE---YTVDVADKGYTLNIRFAGKEK 155
DB 160 LSDFLKSQTP-----AEDTVKSIELAEAKVLANRELDKYQVSDYHKNLINNAKTVEGVK 213
QY 156 TPEEPKEEVTIKANLIYA-DG-----KTQTAEFKGTPEEATAEAYRYAD---LLAKENGK 206
DB 214 DLQAVVESAKKARISEATDGLSDFLKSQTP-----AEDTVKSIELAEAKVLANRELDK 267
QY 207 YTVDVADKGYTLNIRFAGKEKTPPEE-----PKEEVTIKANLIYADKGTQTAEFKGTFAE 260
DB 268 YGVSDYYKNLINNAKTVEGVKALIDEILAALPKTD-TYK---LILNGKTLKGE---TTTE 320
QY 261 A--TAEAYRYADLLAKENGKYTADLEGGYTLNIRFAGKKVDEKPE-----EPMDTY 310
DB 321 AVDAATAEKFVKQYANDG-----VDGEWTDYDDATKFTVTEKPEVIDASELTPAVTTY 374
QY 311 KLILNGKTLKGETTTEAVDAATAEKFVKQYANDGVGDEWTDYDDATKFTVTEKPEVIDA 370
DB 375 KLIVNGKTLKGETTTEAVDAATAEKFVKQYANDGVGDEWTDYDDATKFTVTEKPEVIDA 434
QY 371 SELTPAVTTYKLIVNGKTLKGETTTKAVDAETAETAKFAKQYANDGVGVWYTDATKFTT 430
DB 435 SELTPAVTTYKLIVNGKTLKGETTTKAVDAETAETAKFAKQYANDGVGVWYTDATKFTT 494
QY 431 VTEM 434
DB 495 VTEM 498
RESULT 10
AARL0005
ID AARL0005 standard; protein; 594 AA.
XX
AC AARL0005;
XX
DT 13-MAR-1991 (first entry)
XX
DE Streptococcus GX7805 protein G.
XX
KW Immunoglobulins; Ig.
XX
OS Streptococcus sp GX7805.
XX
FH Key Location/Qualifiers
FT Active-site 304..358
FT /label= Active Site B1
FT Active-site 374..428
FT /label= Active Site B3
FT Active-site 444..498
FT /label= Active Site B2
XX
XX US4977247-A.
XX PN
XX 11-DEC-1990.
XX PD
XX 19-MAY-1989; 89US-0354264.
XX PF
XX 19-MAY-1989; 89US-0354264.
XX PR
XX 14-FEB-1986; 86US-0829354.
XX PR 23-APR-1986; 86US-0854887.
XX PR 17-FEB-1987; 87WO-0500329.
XX PR 19-JUN-1987; 87US-0063959.
XX PR 20-JUN-1988; 88US-0209236.
XX
XX (GENE-) GENEX CORP.


```

FT      /label=B1
FT      Active-site 373..427
FT      /label=B3
FT      Active-site 443..497
FT      /label=B2
XX      US4956296-A.
PN      11-SEP-1990.
XX      20-JUN-1988; 88US-0209236.
XX      20-JUN-1988; 88US-0209236.
PR      14-FEB-1986; 86US-0829354.
PR      23-APR-1986; 86US-0854887.
PR      17-FEB-1987; 87WO-US00329.
PR      19-JUN-1987; 87US-0063959.
XX      (GENE-) GENEX CORP.
XX      Fahnestock SR;
XX      WPI; 1990-297491/39.
XX      N-PSDB; AAQ06019.
XX      Recombinant Protein G variants - obt'd. using a cloned gene
PT      encoding Protein G from Streptococcus sp., used for binding
PT      immunoglobulin.
XX      Disclosure; Fig 9; 48pp; English.
XX      Fragments and variants of the sequence are claimed esp. where
CC      incorporated into a non-pathogenic host eg. E.coli, and expressed
CC      at high levels.
CC      The variants have a higher binding efficiency and capacity for
CC      immunoglobulin, and may be used for purifying, detecting and
CC      isolating antibodies.
XX      Sequence 593 AA;
XX      Query Match 31.0%; Score 692; DB 11; Length 593;
XX      Best Local Similarity 47.9%; Pred. No. 4.6e-33;
XX      Matches 203; Conservative 28; Mismatches 133; Indels 60; Gaps 17;

QY      45 ATSEAYAYADTLKNGEYTDVADKGYTLNIFAGKEKTPPEPKKEVTIKANLIYA-DG 103
DB      101 ADALAKAKADALKEFN-KYGVSDYKNNLINNAKTVEGVKDLQAQVSEAKKARISEATDG 159
QY      104 -----KTQTAEFKGTFFEATAEAYRYADALKKNGE---YTDVADKGYTLNIFAGKEK 155
DB      160 LSDFLKSQTP-----AEDTVKSIELAEAKVLANRELDKYGVSYYKNNLNNAKTVEGVK 213
QY      156 TPEPKKEVTIKANLIYA-DG-----KTQTAEFKGTFFEATAEAYRYAD---LLAKENGK 206
DB      214 DLQAQVSEAKKARISEATDGLSDFLKSQTP-----AEDTVKSIELAEAKVLANRELDK 267
QY      207 YTDVADKGYTLNIFAGKEKTPPEE-----PKKEVTIKANLIYADCKTQTAEFKGTFAE 260
DB      268 YGVSDYKNNLINNAKTVEGVKALIDEILALPKTD-TYK---LILNGKTLKGE---TTTE 320
QY      261 A--TAEAYRYADLLAKENGKRYTADLDGGYTNIRFAGKGVDEKPE-----EPMDTY 310
DB      321 AVDAATAEKVKQYANDG-----VDGEWYDDATKTFVTVEKPEVIDASELTPAVTTY 374
QY      311 KLINGKTLKGETTTEAVDAATAEKVKQYANDGVDEWYDDATKTFVTVEKPEVIDA 370
DB      375 KPVIKNGTLKGETTTEAVDAATAEKVKQYANDGVDEWYDDATKTFVTVEKPEVIDA 434
QY      371 SELTPAVTTYKLVINGKTLKGETTTKAVDAETAETAEKQYANDGVGVWYDDATKTF 430
DB      435 SELTPAVTTYKLVINGKTLKGETTTKAVDAETAETAEKQYANDGVGVWYDDATKTF 494
QY      431 VTEM 434

```

```

DB      495 VTEM 498
||||
RESULT 13
AAP70493
ID      AAP70493 standard; protein; 448 AA.
XX      XX
AC      AAP70493;
XX      06-MAR-1991 (first entry)
XX      Protein G.
XX      Protein G; antibody; Fc receptor;
XX      Streptococcus Lancefield Group G strain.
XX      Key Location/Qualifiers
FH      Active-site 228..297
FT      /label=active site B1
FT      Active-site 298..352
FT      /label=active site B2
XX      WO8705025-A.
XX      27-AUG-1987.
XX      17-FEB-1987; 87WO-US00329.
XX      22-APR-1986; 86US-0854997.
PR      14-FEB-1986; 86US-0829354.
XX      (GENE-) GENEX CORP.
PA      (FAHN/) FAHNESTOCK S R.
XX      Fahnestock S;
XX      WPI; 1987-250197/35.
XX      N-PSDB; AAN70811.
XX      Cloned Protein G gene - used for producing Protein G for
PT      detection and purification of antibodies and treatment of diseases
XX      Disclosure; Fig. 3; 68pp; English.
XX      Protein G expressed by inserting the gene into an expression
CC      vector. A second vector may also be used as a cryptic helper plasmid
CC      to stably maintain the first plasmid in the host cell. Bacterial Fc
CC      receptors such as Protein G can be used to detect and purify
CC      antibodies, and in the treatment of disease. Fc receptors are useful
CC      to purify antibodies to be used in the purification of protein drugs and
CC      as therapeutics. High levels of Protein G can be obtained in conditions
CC      favourable for isolation, using a non-pathogenic host. Suitable
XX      cloning vectors are lambda gt11, M13mp9 and pCX1066.
XX      Sequence 448 AA;
XX      Query Match 30.6%; Score 683; DB 8; Length 448;
XX      Best Local Similarity 53.2%; Pred. No. 1.1e-32;
XX      Matches 176; Conservative 15; Mismatches 92; Indels 48; Gaps 9;

QY      130 KDNGEYTDVADKGYTLNIFAGKEKTPPEPKKEVTIKANLIYAD-CKTQTAEFKGTFFEE 188
DB      45 RNGELT-----NLLGNSETTLALRNEESATADLTAAAVADTVAAAAENAGA 92
QY      189 ATAETARYADLLA-----KENGKYTDVADKGYTLNIFAGKEKTPPEPKKEVTIKAN 241
DB      93 AAWAAAAADALAKADALKEFNKYGVSDYKNNLNNAKTVEGVKDLQAQVSEAKKAR 152
QY      242 LIYA-DG-----KTQTAEFKGTFFEATAEAYRYAD---LLAKENGKRYTADLDGGYTN 292
DB      153 ISEATDGLSDFLKSQTP-----AEDTVKSIELAEAKVLANRELDKYGV-----SDYHKNL 202

```

Db	203	INNACTVEGVKELIDEIALAPKTDYTKLLNGKTLKGETTTTAVDRAATAEKVFKQVAND	262
Qy	344	NGVDGEWTDYDAKTKFTVTTEKPEVIDASELTPAVTTYKLVINGKTLKGETTTTKAVDAETA	403
Db	263	NGVDGEWTDYDAKTKFTVTTEKPEVIDASELTPAVTTYKLVINGKTLKGETTTTKAVDAETA	322
Qy	404	EKAPKQVANDNGVDGVWVWYDDAKTKFTVTTEM	434
Db	323	EKAPKQVANDNGVDGVWVWYDDAKTKFTVTTEM	353
RESULT 15			
AAR07013			
XX	ID	AAR07013 standard; protein; 448 AA.	
XX	AA	AAR07013;	
XX	DT	17-JAN-1991 (first entry)	
XX	DE	Protein G variant with two active sites.	
XX	KW	Immunoglobulin.	
XX	OS	Streptococcus sp. Lancefield Group G strain.	
XX	FH	Key	Location/Qualifiers
FT	FT	Active-site	228..282
FT	FT		/label=B1
FT	FT	Active-site	298..352
FT	FT		/label=B2
XX	PN	US4956296-A.	
XX	XX		
XX	PD	11-SEP-1990.	
XX	XX		
XX	PF	20-JUN-1988;	88US-0209236.
XX	XX		
XX	PR	20-JUN-1988;	88US-0209236.
XX	PR	14-FEB-1986;	86US-0829354.
XX	PR	23-APR-1986;	86US-0854887.
XX	PR	17-FEB-1987;	87WO-US00329.
XX	PR	19-JUN-1987;	87US-0063959.
XX	XX		
PA	(GENE-)	GENEX CORP.	
XX	PI	Fahnestock SR;	
XX	XX		
XX	DR	WPI; 1990-297491/39.	
XX	DR	N-PSDB; AAQ06018.	
XX	XX		
PT	Recombinant Protein G variants - obtd. using a cloned gene		
PT	encoding Protein G from Streptococcus sp., used for binding		
PT	immunoglobulin.		
XX	XX		
PS	Disclosure; Fig 8a-c; 48pp; English.		
XX	XX		
CC	Fragments and variants of the sequence are claimed esp. where		
CC	incorporated into a non-pathogenic host eg. E.coli, and expressed		
CC	at high levels.		
CC	The variants have a higher binding efficiency and capacity for		
CC	immunoglobulin, and may be used for purifying, detecting and		
CC	isolating antibodies.		
XX	XX		
SQ	Sequence	448 AA;	
Query Match 30.6%; Score 683; DB 11; Length 448;			
Best Local Similarity 53.2%; Pred. No. 1.1e-32;			
Matches 176; Conservative 15; Mismatches 92; Indels 48; Gaps			
Qy	130	KDNGEYTVADVADKGYTLINIKFAGEKTEPEPKGEVTIKANLIYAD-GKQTQAEFKGTTEE	188
Db	45	RNGSELT-----NLLGNSETTLALRNEESATADLTAAAVADTVAAAAAENAGA	92

THIS PAGE BLANK (USPTO)

GenCore version 5.1.1.3
Copyright (c) 1993 - 2002 Compugen Ltd.

OM protein - protein search, using sw model

Run on: October 29, 2002, 09:25:41 ; Search time 29,364 Seconds
(without alignments)
2556.865 Million cell updates/sec

Title: US-08-325-278b-3
Perfect score: 2235
Sequence: 1 AVENKEETPETDSEEEV.....GVDGWTYDDATKTFVTVM 434

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5
Searched: 562222 seqs, 172994929 residues
Total number of hits satisfying chosen parameters: 562222

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : SPTREMBL_19:*
1: sp_archaea:*
2: sp_bacteria:*
3: sp_fungi:*
4: sp_human:*
5: sp_invertebrate:*
6: sp_mammal:*
7: sp_mhc:*
8: sp_organalle:*
9: sp_phage:*
10: sp_plant:*
11: sp_rodent:*
12: sp_virus:*
13: sp_vertebrate:*
14: sp_unclassified:*
15: sp_virus:*
16: sp_bacteriap:*
17: sp_archaeap:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	ID	Description
1	2235	100.0	455	2 Q53291	Q53291 streptococc
2	1587.5	71.0	719	2 Q51912	Q51912 peptostrept
3	1263.5	56.5	992	2 Q51918	Q51918 peptostrept
4	706.5	31.6	664	2 Q53975	Q53975 streptococc
5	697.5	31.2	669	2 Q93PM8	Q93PM8 streptococc
6	670	30.0	185	2 Q54181	Q54181 streptococc
7	481	21.5	429	2 Q56212	Q56212 streptococc
8	404	18.1	443	2 Q53974	Q53974 streptococc
9	294	13.2	60	2 Q53337	Q53337 streptococc
10	182.5	8.2	3198	5 Q9UG88	Q9UG88 manduca sex
11	182	8.1	479	10 Q9LF88	Q9LF88 arabidopsis
12	177.5	7.9	2402	2 Q9AER7	Q9AER7 staphylococ
13	175	7.8	448	10 Q9SKP0	Q9SKP0 arabidopsis
14	170.5	7.6	1849	2 Q9SAK2	Q9SAK2 lactobacill
15	170	7.6	2276	2 Q93TY6	Q93TY6 staphylococ
16	168	7.5	448	10 Q96246	Q96246 arabidopsis

17	162.5	7.3	2353	2 P71401	P71401 haemophilus
18	162	7.2	913	13 Q13099	Q13099 xenopus lae
19	161	7.2	2692	4 Q9Y503	Q9Y503 homo sapien
20	160.5	7.2	5458	5 Q90459	Q90459 plasmodium
21	160	7.2	2691	4 Q9NS12	Q9NS12 homo sapien
22	160	7.2	2705	4 Q95303	Q95303 homo sapien
23	160	7.2	2725	4 Q9NYE5	Q9NYE5 homo sapien
24	159	7.1	1856	16 Q97T80	Q97T80 streptococc
25	158	7.1	2045	16 Q9A0K5	Q9A0K5 streptococc
26	157.5	7.0	1166	2 Q86489	Q86489 staphylococ
27	157	7.0	798	13 Q90307	Q90307 carassius a
28	156.5	7.0	564	2 Q69188	Q69188 streptococc
29	156.5	7.0	581	2 Q33742	Q33742 streptococc
30	156.5	7.0	663	2 Q30874	Q30874 streptococc
31	156.5	7.0	701	2 Q9KK48	Q9KK48 streptococc
32	156.5	7.0	969	5 Q9ND19	Q9ND19 plasmodium
33	156	7.0	1368	2 Q93M90	Q93M90 clostridium
34	154.5	6.9	1110	13 Q91255	Q91255 petromyzon
35	154.5	6.9	1822	2 Q07290	Q07290 streptococc
36	154	6.9	737	16 Q9JYK4	Q9JYK4 neisseria m
37	154	6.9	845	5 Q9Y1P8	Q9Y1P8 plasmodium
38	154	6.9	1698	2 Q9LC00	Q9LC00 staphylococ
39	153.5	6.9	492	2 Q56193	Q56193 staphylococ
40	153.5	6.9	573	2 Q56192	Q56192 staphylococ
41	153.5	6.9	701	2 Q9RQT5	Q9RQT5 streptococc
42	153.5	6.9	1873	2 Q9Z4N7	Q9Z4N7 enterococcu
43	153	6.8	897	13 Q13098	Q13098 xenopus lae
44	152	6.8	3147	5 Q17464	Q17464 caenorhabdi
45	151.5	6.8	717	2 Q9EY85	Q9EY85 clostridium

ALIGNMENTS

RESULT 1

Q53291 PRELIMINARY; PRT; 455 AA.
ID Q53291
AC Q53291:
DT 01-NOV-1996 (TRENBLrel. 01, Created)
DT 01-NOV-1996 (TRENBLrel. 01, Last sequence update)
DT 01-DEC-2001 (TRENBLrel. 19, Last annotation update)
DE PROTEIN LG (FRAGMENT).
OS Streptococcus sp.
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Streptococcaceae;
OC Streptococcus.
OX NCBI_TaxID=1306;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=93094283; PubMed=1460053;
RA Kihlberg B.M., Sjoberg U., Kastern W., Bjorck L.;
RT Protein LG: a hybrid molecule with unique immunoglobulin binding properties.;
RL J. Biol. Chem. 267:25583-25588(1992).
DR EMBL; S50809; AAA03280.1; ..
DR HSSP; P06654; 1PGX.
DR InterPro; IPR003147; B1.
DR InterPro; IPR000724; IGG_bind_B.
DR Pfam; PF02246; B1; 4.
DR Pfam; PF01378; IGG_binding_B; 2.
FT NON_TER 455 455
SQ SEQUENCE 455 AA; 49926 MW; 381FC235BBC8307B CRC64;

Query Match 100.0%; Score 2235; DB 2; Length 455;
Best Local Similarity 100.0%; Pred. No. 1e-100;
Matches 434; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 AVENKEETPETDSEEEVITKANLIFANGSTQTEFFKTFEKATSEAYAYADTLKKDN 60
DB 22 AVENKEETPETDSEEEVITKANLIFANGSTQTEFFKTFEKATSEAYAYADTLKKDN 81
QY 61 GETYVDVADKGYTLNFKFAGKKTPEEPKEVITKANLIYADGKTQTAEPKGFEEATAE 120
DB 82 GETYVDVADKGYTLNFKFAGKKTPEEPKEVITKANLIYADGKTQTAEPKGFEEATAE 141

THIS PAGE BLANK (USPTO)

QY 121 AYRYADALKKDNGEYTVDVADKGYTLNIRFAGKEKTPPEPKKEVTIKANLIYADGKTQTA 180
|||||
Db 142 AYRYADALKKDNGEYTVDVADKGYTLNIRFAGKEKTPPEPKKEVTIKANLIYADGKTQTA 201
|||||
QY 181 EFKGTEETAEAYRYADLLAKENGKYYTVDVADKGYTLNIRFAGKEKTPPEPKKEVTIKA 240
|||||
Db 202 EFKGTEETAEAYRYADLLAKENGKYYTVDVADKGYTLNIRFAGKEKTPPEPKKEVTIKA 261
|||||
QY 241 NLTYADGKTQTAEFKGTFAEATAEAYRYADLLAKENGKYYTVADLEGGYTNIRFAGKYYD 300
|||||
Db 262 NLTYADGKTQTAEFKGTFAEATAEAYRYADLLAKENGKYYTVADLEGGYTNIRFAGKYYD 321
|||||
QY 301 EKPEEPMDYTKLILNGKTKLGETTFAVDAATAEAKVFKOYANDNGVDGEMTYDDATKTF 360
|||||
Db 322 EKPEEPMDYTKLILNGKTKLGETTFAVDAATAEAKVFKOYANDNGVDGEMTYDDATKTF 381
|||||
QY 361 VTEKPEVIDASELTPAVTYTKLVINGKTKLGETTTKAVDAETAETAKAFKYOYANDNGVDGYW 420
|||||
Db 382 VTEKPEVIDASELTPAVTYTKLVINGKTKLGETTTKAVDAETAETAKAFKYOYANDNGVDGYW 441
|||||
QY 421 TYDDATKTFVTTEM 434
|||||
Db 442 TYDDATKTFVTTEM 455
|||||

RESULT 2

Q51912 PRELIMINARY; PRT; 719 AA.
ID Q51912
AC Q51912;
DT 01-NOV-1996 (TREMBlrel. 01, Created)
DT 01-NOV-1996 (TREMBlrel. 01, Last sequence update)
DT 01-DEC-2001 (TREMBlrel. 19, Last annotation update)
DE PROTEIN L PRECURSOR.
OS Peptostreptococcus magnus.
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Clostridiaceae;
OC Finegoldia.
OX NCBI_TaxID=1260;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=312;
RX MEDLINE=90215984; PubMed=2108927;
RA Kastern W., Holst E., Nielsen E., Sjoerling U., Bjorck L.;
RT "Protein L, a bacterial immunoglobulin-binding protein and possible
RL virulence determinant";
RN Infect. Immun. 58:1217-1222(1990).
[2]
RP SEQUENCE FROM N.A.
RC STRAIN=312;
RX MEDLINE=92316971; PubMed=1618782;
RA Bjoerck L., Sjoerling U., Kastern W.;
RT "Structure of peptostreptococcal protein L and identification of
RT repeated immunoglobulin light chain-binding domain";
RL J. Biol. Chem. 267:12820-12825(1992).
DR EMBL: M86697; AAA25612.1;
DR InterPro: IPR003147; B1.
DR InterPro: IPR001899; Gram_pos_anchor.
DR Pfam: PF02246; B1; 5.
DR Pfam: PF00746; Gram_pos_anchor; 1.
KW Signal.
FT SIGNAL.
FT CHAIN 1 18 POTENTIAL.
FT CHAIN 19 719 PROTEIN L.
SQ SEQUENCE 719 AA; 78983 MW; 963A8D76D5E34DD2 CRC64;

Query Match 71.0%; Score 1587.5; DB 2; Length 719;
Best Local Similarity 76.3%; Pred. No. 3e-69; Indels 35; Gaps 6;
Matches 334; Conservative 13; Mismatches 56;

QY 2 VENKEETPTPTDSEEEVTIKANLIFANGSTQTAFFKGTFAEATSEAYAYADTLKKDNG 61
|||||
Db 95 VENKEETPTPTDSEEEVTIKANLIFANGSTQTAFFKGTFAEATSEAYAYADTLKKDNG 154
|||||
QY 62 EYTVADKGYTLNIRFAGKEKTPPEPKKEVTIKANLIYADGKTQTAFFKGTFAEATAE 121
|||||

Db 155 EYTVADKGYTLNIRFAGKEKTPPEPKKEVTIKANLIYADGKTQTAFFKGTFAEATAE 214
|||||
QY 122 YRYADALKKDNGEYTVDVADKGYTLNIRFAGKEKTPPEPKKEVTIKANLIYADGKTQTA 181
|||||
Db 215 YRYADALKKDNGEYTVDVADKGYTLNIRFAGKEKTPPEPKKEVTIKANLIYADGKTQTA 274
|||||
QY 182 FKGTETAEATAEAYRYADLLAKENGKYYTVDVADKGYTLNIRFAGKEKTPPEPKKEVTIKAN 241
|||||
Db 275 FKGTETAEATAEAYRYADLLAKENGKYYTVDVADKGYTLNIRFAGKEKTPPEPKKEVTIKAN 334
|||||
QY 242 LIYADGKTQTAFFKGTFAEATAEAYRYADLLAKENGKYYTVADLEGGYTNIRFAGKYYD 301
|||||
Db 335 LIYADGKTQTAFFKGTFAEATAEAYRYADLLAKENGKYYTVADLEGGYTNIRFAGKYYD 394
|||||
QY 302 KPEEPMYD-TYK-LILNGKTKLGETTFAVDAATAEAKVFKOYANDNGVD-GEMTYDDATK 357
|||||
Db 395 KPEEKEQVTKENIYFEDGTGTATFKGTFAEATAEAY--RYADLLSKERHGYTADLEGG 452
|||||
QY 358 TPTVT-----EKPEVIDASELTPAVTYTKLVINGKTKLGETTTKAVDAETAETAE 404
|||||
Db 453 GVTINIRFAGKEKTPPEPKKEVQD-----GVASYEEAAEAAKEALKND 496
|||||
QY 405 KAFKYOYANDNGVDGEMTY 422
|||||
Db 497 DVNKSYYTIROGADGRYY 514
|||||

RESULT 3

Q51918 PRELIMINARY; PRT; 992 AA.
ID Q51918
AC Q51918;
DT 01-NOV-1996 (TREMBlrel. 01, Created)
DT 01-NOV-1996 (TREMBlrel. 01, Last sequence update)
DT 01-DEC-2001 (TREMBlrel. 19, Last annotation update)
DE PROTEIN L PRECURSOR.
OS Peptostreptococcus magnus.
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Clostridiaceae;
OC Finegoldia.
OX NCBI_TaxID=1260;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=3316;
RX MEDLINE=95078460; PubMed=7987012;
RA Murphy J.P., Troewen A.R., Duggleby C.J.;
RT "Nucleotide sequence of the gene for peptostreptococcal protein L";
RL DNA Seq. 4:259-265(1994).
DR EMBL: L04466; AAA67503.1;
DR HSP; Q51911; IGAB.
DR InterPro: IPR003147; B1.
DR InterPro: IPR002988; GA.
DR InterPro: IPR001899; Gram_pos_anchor.
DR Pfam: PF02246; B1; 4.
DR Pfam: PF01468; GA; 4.
DR Pfam: PF00746; Gram_pos_anchor; 1.
KW Signal.
FT SIGNAL.
FT CHAIN 1 24 POTENTIAL.
FT CHAIN 25 992 PROTEIN L.
SQ SEQUENCE 992 AA; 108700 MW; 9CFF5771578A5DCE CRC64;

Query Match 56.5%; Score 1263.5; DB 2; Length 992;
Best Local Similarity 58.8%; Pred. No. 1.8e-53;
Matches 281; Conservative 36; Mismatches 90; Indels 71; Gaps 12;

QY 2 VENKEETPTPTDSEEEVTIKANLIFANGSTQTAFFKGTFAEATSEAYAYADTLKKDNG 61
|||||
Db 241 MERKUSEKETPE--PEEEVTIKANLIFADGSTQNAEAFKGTFAKAVSDAYAYADLKKDNG 298
|||||
QY 62 EYTVADKGYTLNIRFAGKEKTPPEPKKEVTIKANLIYADGKTQTAFFKGTFAEATAE 121
|||||
Db 299 EYTVADKGLNIRFAGKEKTPPEPKKEVTIKANLIFADGKTQTAFFKGTFAEATAE 358
|||||
QY 122 YRYADALKKDNGEYTVDVADKGYTLNIRFAGKE--KTPEEPKEVTIKANLIYADGKTOT 179
|||||

THIS PAGE BLANK (USPTO)

GenCore version 5.1.3
Copyright (c) 1993 - 2002 Compugen Ltd.

OM protein - protein search, using sw model

Run on: October 29, 2002, 09:25:41; Search time 20.636 Seconds

(without alignments)
2556.865 Million cell updates/sec

Title: US-08-325-278B-1

Perfect score: 1565
Sequence: 1 AVENKEETPEPTDSEEV.....GGYTINIRAGKRVDEKPEE 305

Scoring table: BIOSUM62

Gapop 10.0, Gapext 0.5

Searched: 562222 seqs, 172994929 residues

Total number of hits satisfying chosen parameters: 562222

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database: SPTREMBL_19:*

- 1: sp.archaea:*
- 2: sp.bacteria:*
- 3: sp.fungi:*
- 4: sp.human:*
- 5: sp.invertebrate:*
- 6: sp.mammal:*
- 7: sp.mhc:*
- 8: sp.organelle:*
- 9: sp.phage:*
- 10: sp.plant:*
- 11: sp.podent:*
- 12: sp.virus:*
- 13: sp.vertebrate:*
- 14: sp.unclassified:*
- 15: sp.virus:*
- 16: sp.bacteriaph:*
- 17: sp.archaeap:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1565	100.0	455	2	053291 streptococc
2	1561	99.7	719	2	051912 peptostrept
3	1226.5	78.4	992	2	051918 peptostrept
4	167	10.7	664	2	053975 streptococc
5	166	10.6	479	10	091F88 arbidopsi
6	166	10.6	669	2	093EM8 streptococc
7	151	9.6	448	10	09SKP0 arbidopsi
8	149	9.5	448	10	096246 arbidopsi
9	148.5	9.5	2276	2	093WY6 streptococc
10	141.5	9.0	1110	13	091255 petromycon
11	141	9.0	1849	2	09S4K2 lactobacill
12	137.5	8.8	992	16	09WZ46 thermotoga
13	137.5	8.8	5458	5	09U459 plasmodium
14	136	8.7	506	2	053837 salmonella
15	134.5	8.6	3198	5	09U8G8 manduca sex
16	132.5	8.5	737	16	09JYK4 neisseria m

17	132.5	8.5	17352	5	095YM2 procambarus
18	131.5	8.4	891	2	004111 enterococcu
19	131.5	8.4	1166	2	086489 staphylococ
20	130.5	8.3	564	2	069188 streptococc
21	130.5	8.3	581	2	033742 streptococc
22	130.5	8.3	663	2	030874 streptococc
23	130.5	8.3	701	2	09KK48 streptococc
24	130.5	8.3	1141	16	099W46 staphylococ
25	130.5	8.3	1733	2	09K114 staphylococ
26	129.5	8.3	1171	2	09KMX6 staphylococ
27	129	8.2	938	16	0927R4 staphylococ
28	128.5	8.2	1119	13	P87344 theragra ch
29	128.5	8.2	1035	5	021380 caenorhabd
30	127.5	8.1	1035	5	021380 caenorhabd
31	127.5	8.1	701	2	09B0T5 glycine max
32	127.5	8.1	728	2	09EY85 mycoplasma
33	127.5	8.1	843	2	047802 enterococc
34	127	8.1	913	13	013099 xenopus lae
35	126.5	8.1	1035	5	021380 caenorhabd
36	126.5	8.1	2045	16	09AOK5 streptococ
37	126	8.1	955	2	045574 bacillus sp
38	125.5	8.0	673	10	02M9S5 streptococ
39	125.5	8.0	1510	5	025920 plasmodium
40	125	8.0	463	10	039871 glycine max
41	125	8.0	2083	5	09N435 caenorhabd
42	124.5	8.0	1302	2	049547 mycoplasma
43	124	7.9	495	16	099W46 staphylococ
44	124	7.9	506	2	03834 salmonella
45	124	7.9	1489	10	096449 phytophthor

ALIGNMENTS

RESULT 1

ID 053291 PRELIMINARY: PRT: 455 AA.

AC 053291;
 DT 01-NOV-1996 (TREMBLrel. 01, Created)
 DT 01-NOV-1996 (TREMBLrel. 01, Last sequence update)
 DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)
 DE PROTEIN IG (FRAGMENT).
 OS Streptococcus sp.
 OC Bacteria; Firmicutes; Bacillus/Clostridium group; Streptococcaceae;
 OC Streptococcus
 OX NCBI_TaxID=1306;
 RN (1)
 RP SEQUENCE FROM N.A.
 RX MEDLINE-93094283; PubMed-1460053;
 RA Kihlberg B.M., Sjobring U., Kasten W., Bjorck L.,
 RT "Protein IG: a hybrid molecule with unique immunoglobulin binding
 properties.";
 RT J. Biol. Chem. 267:25583-25588(1992).
 DR EMBL: S50809; AAA03280.1;
 DR HSSP: P06654; IPGX.
 DR InterPro: IPR003147; B1.
 DR InterPro: IPR00724; Igg_bind_B.
 DR Pfam: PF02246; B1; 4.
 DR Pfam: PF01378; Igg_binding_B; 2.
 FT NON_TER 455
 SQ SEQUENCE 455 AA; 49926 MW; 381FC235B8C8307B CRC64;

Query Match 100.0%; Score 1565; DB 2; Length 455;
 Best Local Similarity 100.0%; Pred. No. 9.1e-79;
 Matches 305; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY	1	AVENKEETPEPTDSEEVITKANLIFANGSTOAEKGFPEKTSAYAYADLKDN	60
DB	22	AVENKEETPEPTDSEEVITKANLIFANGSTOAEKGFPEKTSAYAYADLKDN	81
QY	61	GEYTVADKGYTLNIRKFKGKKEPEPEKEEYTIKANLIVADGKTQTAFFGTFPEATAE	120
DB	82	GEYTVADKGYTLNIRKFKGKKEPEPEKEEYTIKANLIVADGKTQTAFFGTFPEATAE	141

Db	Qy	242	LIYADGKQTQTEFEKGTFAEATAEAYRYADLLAKNGKGYTADLEGGYTINIRFAGKRVDE	301
Db	335	LIYADGKQTQTEFEKGTFAEATAEAYRYADLLAKNGKGYTADLEGGYTINIRFAGKRVDE	394	
Qy	302	KPEE 305		
Db	395	KPEE 398		
RESULT 3				
ID	051918	PRELIMINARY;	PRT;	992 AA.
AC	051918;			
DT	01-NOV-1996 (TrEMBLrel. 01, Created)			
DT	01-NOV-1996 (TrEMBLrel. 01, Last sequence update)			
DT	01-DEC-2001 (TrEMBLrel. 19, Last annotation update)			
DE	PROTEIN L PRECOURSOR.			
OS	Peptostreptococcus magnus.			
OC	Bacteria; Firmicutes; Bacillus/Clostridium group; Clostridiaceae;			
CC	Flagellidia.			
OX	NCBI_Taxid=1260;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RC	STRAIN=3316;			
FX	MEDLINE=95078460; PubMed=7987012;			
RT	Murphy J.P., Townen A.R., Dugdaley C.J.;			
RA	"Nucleotide sequence of the gene for peptostreptococcal protein L.";			
RL	DNA Seq. 4:259-265(1994).			
DR	EMBL, L04466; AAA67503.1; ..			
DR	HSSP; Q51911; IGAB.			
DR	InterPro; IPR003147; B1.			
DR	InterPro; IPR002988; GA.			
DR	InterPro; IPR001899; Gram_pos_anchor.			
DR	Pfam; PF02246; B1; 4.			
DR	Pfam; PF01468; GA; 4.			
KW	Pfam; PF00746; Gram_pos_anchor, 1.			
DR	Signal.			
FT	SIGNAL	1	24	POTENTIAL.
FT	CHAIN	25	992	PROTEIN L.
SQ	SEQUENCE	992 AA;	108700 MW;	9CF5F771578A5DCE CRC64;
Qy	2	VENKEETPEPTESEEVITKANLIRANSTQTAERKGFSEKATSAVAYADLLKNDG	61	
Db	241	MERTLSKEIPE--PEEEVITKANLIRADSTQNAERKGFPAAVASDAVYADLLKNDG	298	
Qy	62	EYTVADVADKGYTINIRFAGKEKTPPEEKEEVTIKANLIVADGKTQTAEFKGTPEEATAEA	121	
Db	299	EYTVADVADKGYTINIRFAGKEKTPPEEKEEVTIKANLIVADGKTQTAEFKGTPEEATAEA	358	
Qy	122	YRYADALKRNGEYTVADVADKGYTINIRFAGKE--KTPEEKEEVTIKANLIVADGKTQ	179	
Db	359	YAYADLLAKENGEGYTDLEGGNTINIRFAGKETPEEPEEKEEVIKVNLIADGKIQT	418	
Qy	180	AEEFGTPEEATAEAYRYADLLAKENGKGYTVADVADKGYTINIRFAGKE--KTPEEKEEVT	237	
Db	419	AEEFGTPEEATAEAYRYADLLAKENGKGYTVADVADKGYTINIRFAGKETPEEPEEATAEA	478	
Qy	238	IKANLIVADGKTQTAERKGFPAEATAEAYRYADLLAKENGKGYTADLEGGYTINIRFAGK	297	
Db	479	IKVNLIADGKTQTAERKGFSEKATSAEAYRYADLLAKVNGEYTDLEGGYTINIRFAGK	538	
Qy	298	KVDEKPEE 305		
Db	539	---EQPEE 543		